The Influence of Gestures on Variability of Human Cognition during the Gear-System Task

Wilhelmina H. M. Bakker

S4512316

Department of Psychology, University of Groningen

PSB3N-BT15: Bachelor Thesis

Group Number: 2324_1a_09a

Supervisor: dr. L. de Jonge-Hoekstra

Second evaluator: prof. dr. E.H. Gordijn

In collaboration with: M. Brink, F.C.T. Burg, S. Das, G.B Selles and A.E. Wachtmeester

January 11, 2024

Abstract

Previous research has found that gestures influence adopted strategies during problem solving of the gear-system task. When gestures are used, a perceptual-motor strategy is usually adopted, whereas when gestures are restricted, a more abstract strategy is used. Next to that, strategy switches may indicate differences in variability from a functional dynamic perspective. The present study aims to replicate previous studies about these topics and analyze them further. Two hypotheses were evaluated about 1) the effect of gesture use (restricted or allowed) on discovering an abstract strategy during the gear-system task and 2) how greater variability will be shown in the earlier trials of the experiment, when gesture use is restricted. Furthermore, variability and strategy switches will be visualized to exploratively investigate the potential influx of variability that may occur before strategy switching. Participants (*N = 120*) were randomly assigned to two conditions, gesture-restricted and gesture-allowed, and asked to solve the gear-system task. Video recordings were made and through coding and multimodal processing, including recurrence quantification analysis, a large dataset was produced to investigate the hypotheses. Our study did not find an effect of condition (gesture-restricted or gesture-allowed) on using a more abstract strategy. Next to that variability was not larger in the first trials for the gesture-restricted condition. The visual comparison, however, gave a promising insight into variability and strategy switching. Future research should focus on further replicating gesture studies and investigating variability with strategy changes to gain more insight in human cognition.

Keywords: Gestures, problem solving, strategy use, functional dynamic theory, recurrence quantification analysis

The Influence of Gestures on Variability of Human Cognition during the Gear-System Task

Gestures are part of our everyday communication. This can be communication with others as well as communication with oneself (Zurina & Williams, 2011). Counting, for example, is a gesture that can help one keep track of something or solve a simple mathematical problem. In the domain of problem solving, gestures can be beneficial in representing the problem at hand to more easily come to a solution (Chu & Kita, 2011). Next to that, research has shown that gesture use may influence the strategies one employs when solving a problem (Alibali et al., 2011). More specifically, Alibali et al. (2011) investigated the *gear-system task* in which one must predict directions of gears in a sequence and found that one way of solving this task is by representing these gears with gestures. However, they also found that after the discovery of another strategy, gestures can become superfluous. This paper will further investigate the relationship between strategies and gesture use on the gear-system task. Moreover, it will apply a functional dynamic systems' perspective on this relationship.

Gestures

Gestures serve a multitude of functions to speakers as well as listeners (Wagner et al., 2014). Examples of these functions can be describing a picture or solving a specific task, as mentioned before. Gestures can be defined as body movements that carry meaning and that are rooted in embodied, dynamic, and intersubjective experiences (Will, 2021). Because gestures can be used in many ways, it is difficult to divide them into mutually exclusive categories. Therefore, McNeill (2005) has created four continua on which gestures can be placed (Wagner et al., 2014). The first continuum is their *relationship to speech*, in which obligatory presence of and absence of gestures during speech are at the two far ends of the spectrum. Secondly, gestures can be ranged on their *relationship to linguistic properties*. For example, waving to someone is a way of saying "hello" and would therefore be high in linguistic properties. Next is the continuum of the *relationship to conventions*. A gesture is placed on how well it is known to the general population. A very conventional gesture in Western culture for instance is shaking your head from side to side. This is known as a gesture for disagreement (Bross, 2020). And lastly, according to McNeill (2005), gestures can be ranged on their *character of semiosis*. This continuum represents the extent to which gestures represent objects or events. For example, moving one's hands meaninglessly during a conversation is a gesture that is low in semiosis, it can be seen as global and does not represent a certain object or image.

Solving the Gear-System Task with Gestures

This last domain of semiosis is relevant for gestures that are used during problem solving, more specifically the gear-system task (Arzarello et al., 2009). During this task, a participant is asked to imagine gears that are arranged in a sequence. The participant is asked to predict in what direction a certain gear turns after hearing in what direction another gear is turning, e.g. "If gear 4 is turning clockwise, in what direction is gear 7 turning?". Most participants start solving these tasks by tracing the forces of the gears (Stephen & Dixon, 2009). This can be done with hand gestures or pointing one's finger and turning it in the imagined gears direction. This particular way of solving the gear problem is a *perceptual motor strategy* (Alibali et al., 2011), in which gestures are used to visualize and solve the problem at hand. These gestures are high in semiosis as they are used to represent and analyze the gears from the task (Arzarello et al., 2009).

Solving the Gear-System Task without Gestures

Research has shown that allowing or prohibiting people to use their hands in the gear-system task could influence the cognitive strategies that they adopted to solve a problem (Alibali et al., 2011). More specifically, Alibali et al. (2011) found that participants were likely to use a perceptual-motor strategy when they used gestures to form a representation of the turning gears. However, when they could not use gestures, an abstract strategy was more often adopted. This is the so-called *parity strategy* (Alibali et al., 2011). This strategy is a more abstract way of problem solving in which the participant notices that odd numbers of gears turn in the same direction as the first gear, and even numbers of gears turn in the opposite direction. For example, when the first gear turns clockwise, it makes the following even gear turn counterclockwise which makes the next uneven gear turn clockwise again and so forth.

Dynamic Systems Theory: Self-Organisation, Dissipative Structures, Entropy & Attractors

The perceptual motor and parity strategy are quite different ways of solving the gear problem. Nonetheless, participants tended to switch from the perceptual-motor strategy to the more abstract parity strategy after finishing some trials (Alibali et al., 2011). *Dynamic systems theory* gives an explanation for this phenomenon. This theory is based on thermodynamics and can be applied to human cognition (De Bari & Dixon, 2022; Stephen & Dixon, 2009; Stephen, Dixon & Isenhower, 2009). Research in this field has focussed on how problem-solving works and, more specifically, what mechanisms are behind going from one strategy to another, or the 'ah-moment.' The occurrence of a new strategy can be seen as the emergence of cognitive structure (Stephen & Dixon, 2009). When this is applied to the gear-system task, this means that the parity strategy self-organizes in the cognitive system after some use of the perceptual motor strategy.

The *self-organization* of a new strategy is driven by energy flow, according to dynamic systems theory (Stephen & Dixon, 2009). First it is important to note that the human cognitive system is similar to a *dissipative structure* from thermodynamics (De Bari & Dixon, 2022). These structures are self-organized systems that are driven by flows of energy. When enough energy is present, the system can spontaneously create new structures. This energy comes in from the surroundings, therefore these systems fundamentally rely on their environment (Kondepudi et al., 2020). Next to this, these systems respond to environmental changes in an adaptive way. Some of the energy from the environment is absorbed and used for tasks in the system, but some of the energy is not. This unusable energy is called *entropy* (Schrödinger, 1944) and the build-up of this is what causes the system to destabilize, increasing its variability (Stephen & Dixon, 2009). To increase the stability of the system again, it can offload excessive energy back to the environment. During this, self-organization of structure takes place and global patterns emerge. These are *attractors*, stable states that the system often returns to (de Jonge-Hoekstra, 2021).

Dynamic Systems Theory and the Gear-System Task

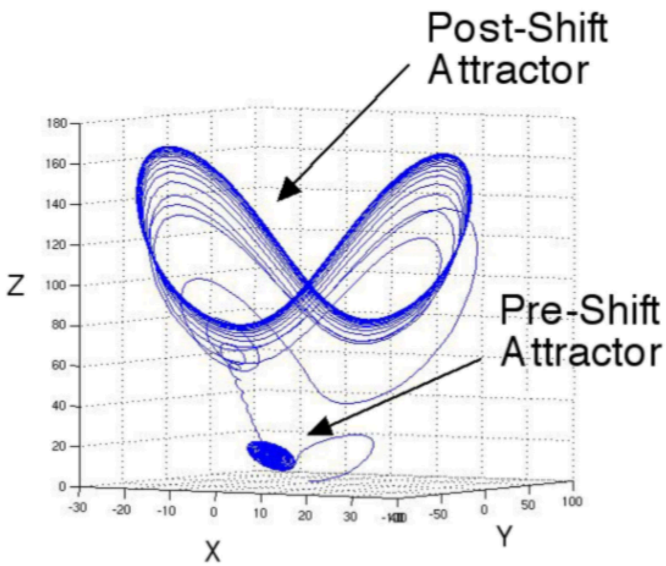
In the gear system task, the different strategies can be seen as different attractor states. As mentioned before, it has been found that most individuals start with applying perceptual-motor strategies and later switch to the more abstract parity strategies (Alibali et al., 2011). In research by Stephen, Dixon & Isenhower (2009) it has been found that during the switch from a perceptual motor to an abstract strategy, there is an increase in variability. The current study hypothesizes and aims to measure what this implies when it is applied to the gear-system task with the gesture-allowed and gesture-restricted condition. The gesture-restricted condition restricts the usage of perceptual-motor strategies (Alibali et al., 2011), which may increase the variability in the system. As hand-movements are restricted, the system is in need of a solution to the gear-system task that does not require gestures. Next to that, perceptual-motor strategies are time-consuming, which may also increase the system's variability as it aims for efficiency (Stephen & Dixon, 2009). The latter can occur in both the gesture-allowed and gesture-restricted condition. Self-organization then takes place as a new attractor state of the parity strategy emerges, and the cognitive system becomes more efficient in solving the gear-system task.

Recurrence Quantification Analysis

To investigate how self-organization takes place, it is essential to measure the stability of the system that employs the strategies during problem solving. Variability of a system can be measured through recurrence quantification analysis (RQA) (Webber & Zbilut, 2005). The attractor states are places on a *phase-space trajectory*, a complex plane on which the time series of a system is plotted against a lagged version of itself. An example of such a phase space trajectory can be viewed in Figure 1 (Stephen & Dixon, 2009). Through autocorrelation the recurrence of a system into its own states on the phase space trajectory can be measured (Stephen, Dixon & Isenhower, 2009). One specific factor that results from this analysis is the *recurrence rate* of a variable. The recurrence rate is the percentage that a system recurs to its previous states during a given period (Webber & Zbilut, 2005). When the recurrence rate is low, the system is in a state of high variability, not recurring to its previous states. A decrease of the recurrence rate could thus indicate that self-organization may take place.

Figure 1

An Example of a Phase-Space Trajectory



Note. This figure shows the phase-space trajectory of the Lorentz system. This is a system that can switch from one attractor state to another. As shown in the figure, it switches from the pre-shift attractor to the post-shift attractor. From "The self-organization of insight: Entropy and power laws in problem solving" by Stephen, D. G., & Dixon, J. A., 2009, *Journal of Problem Solving*, *2*(1), p. 83.

Current Study

The current study aims to gain insight in how gestures might interact with variability of the cognitive system during problem solving. Our main research question is: "*How does the use, or non-use of hand movements during problem solving influence variability of the cognitive system and the occurrence of self-organization and switching to the parity strategy?*" Participants are divided into a gesture-allowed and gesture-restricted condition. In the gesture-restricted condition, participants are unable to move their hands, whereas in the gesture-allowed condition they can do so. RQA will be applied to motion data to measure the recurrence rates of participants' body movements during the gear-system task. The increase or decrease in recurrence rates will be calculated from trial to trial per participant to investigate whether this is related to the condition the participant was in and/or to the moment of switching from perceptual motor to the parity strategy.

We hypothesize that the non-gesture group will more often apply the parity strategy than the gesture group (hypothesis 1). This would be a direct replication of Alibali et al. (2011). Moreover, we hypothesize that participants in the gesture-restricted condition have lower recurrence rates in the earlier stages of the experiment compared to participants in the gesture-allowed condition (hypothesis 2). They are expected to have greater variability because they are not able to use the perceptual motor strategy through gesturing to solve the problem. Experimentally, mean recurrence rates per trial will be compared to the number of participants that switch to the parity strategy per trial. This will be done per condition to explore whether there is a trend in self-organization, or the discovery of the parity strategy, and whether this differs for participants in the gesture-allowed or gesture-restricted condition.

Methods

Participants

The recruitment of participants was done using a convenience sampling approach, wherein students at a bachelor level affiliated with the faculty of Behavioral and Social Sciences were included in the study, as in Alibali et al. (2011). Participants were invited to partake in the study through the experiment leaders' personal network. A total of 120 students participated in our study. Of this sample, 52 were excluded (see Table 1 for an overview of the reasons for exclusion). According to the protocol as described by Alibali and colleagues (2011), all participants should be offered a second chance when the first answer was incorrect. In our case, we also included participants who were given the correct answer instead of a second chance. The reasoning behind this was that we expect the influence of this deviation from Alibali et al.'s (2011) protocol on the discovery of the parity strategy to be minimal.

The final sample consisted of 68 participants (n_M = 9, n_F = 59). The participants' ages ranged from 19 years old to 25 years old (*M* = 20.72; *SD* = 1.52). All of the included participants were Dutch. 35 participants were assigned to the gesture-allowed condition, and 33 participants were assigned to the gesture-restricted condition. All participants signed the informed consent form. This experiment was approved by the Ethics Committee of the Psychology department of the Faculty of Behavioral and Social Sciences, University of Groningen (codes: PSY-1819-S-0037 and PSY-17209-O).

Table 1

Number of Participants Excluded per Exclusion Criteria

Exclusion Criteria	Number of Participants Excluded
Coding was not available	18
Protocol was not followed	16
Participant not completely visible	11
Hands constraint but still moved them	3
ELAN error	2
Video was incomplete	1
No understanding of what a gear was	1
Total	52

Note. This table consists of the number of participants that were excluded per exclusion criteria followed by the total number of participants that were excluded.

Materials

The experiment took place in the research room of the host faculty. Participants were asked to fill out a questionnaire on Qualtrics. They were randomly assigned to one of the two conditions by flipping a coin. In the gesture-restricted condition, a board with oven mitts or gloves that were connected to two cans via Velcro was used to prevent participants from moving their hands. In the gesture-allowed condition, a board was used to strap the participant's feet using Velcro, preventing them from moving their feet. This was done to provide a level of strangeness, similar to the gesture-restricted condition. A screen divided the experiment leader from the participant, so that they could not see each other. Two cameras were used to record the participant from different angles. The recordings were saved on the servers of the University of Groningen. The best videos were uploaded for further analysis. All experimenters followed the same experiment protocol. To analyze the data, we used JASP.

To process the data from the videos, we used multiple programs. First, we used OpenShot Video Editor to trim the videos and separately save the video, as well as the audio and the image sequence of the video (100 Hz). Audacity was used to remove the background noise from the audio and R Studio 2023.03.0 to extract the amplitude envelope (Pouw & Trujillo, 2019). Then, we used MATLAB R2022b to extract pixel change between frames to extract the overall body movement intensity from the videos (Paxton & Dale, 2012). Lastly, we used ELAN 6.4 to code the speech, trials, and strategies of the participants. The experiment protocol and coding manual was based on Alibali et al. (2011). The instruction manual and further scripts for the data processing and coding were provided by the thesis supervisor. Details about the coding can be found in the Data Processing section.

Procedure

The experiment was conducted in Dutch. Two experimenters were present: One experiment leader conducted the study and one assisted. After being welcomed into the room, participants were handed the informed consent form, which they then signed. Participants were told the research is about solving gear problems and that they would be recorded. Unlike in Alibali et al.'s (2011) study, the camera was not hidden. Participants were then asked to fill out a short questionnaire on Qualtrics about their age, gender, study, and nationality. After this, participants were asked to sit down in a chair and the experiment leader constrained their feet by strapping them to a board (gesture-allowed condition), or constrained their hands by having participants wear gloves or oven mitts (gesture-restricted condition). After this, the experiment leader stepped behind the screen and started the experiment.

The participants were instructed to solve six sets of gear problems and were asked to think out loud. The gear problems were presented in the same order for all the participants. The first problem was: "Imagine four gears connected in a horizontal line. If you turned the gear on the left clockwise, what would the gear on the right do?" The same question was asked about seven, nine, five, eight, and six gears, respectively. In case a participant provided an incorrect answer, they were given a second opportunity. If the answer was still incorrect, the experiment leader would move on to the next trial. After the sixth trial, the experiment leader announced that the experiment was over.

Data Processing

First, we trimmed the videos to the length of the actual experiment using OpenShot Video Editor. The videos were saved as .mp4. Using the same program, we saved the audio file from the trimmed video as .mp3, as well as the image sequence of all the video frames as .png. Next, the background noise of the .mp3 audio was removed using the program Audacity, saving the file as .wav. After this, we extracted the amplitude envelope, which contains the changes in the amplitude of sound over time of the .wav audio files using a script (Pouw & Trujillo, 2019) in R Studio 2023.03.0. Furthermore, the amplitude envelope of participants' speech was processed in R Studio to automatically determine when someone in the video started or stopped talking, as accurately as possible. Then, body movements from the videos were extracted using the Frame Differencing Method in MathWorks Matlab MATLAB R2022b (Paxton & Dale, 2012). Frame differencing is the process of subtracting two frames from each other to obtain the pixel difference; the subtraction is typically done pixel by pixel for pixels in the image with the same coordinates. We focused on upper body movements, i.e.,. head movements and hand gestures. These frame differences were extracted and saved as a .csv file.

Using ELAN, we coded the videos. First, speech was linked corresponding to the person speaking, i.e., it was coded when either the participant or the experiment leader spoke, based on the automatically identified start and stop times of speech in R. After this, we sectioned the six trials in the videos out and coded them individually. Strategies were coded at two levels: General and within trial. The general strategies used by the participants were coded as the overall strategy of a trial, which is similar to how Alibali et al. (2011) coded participants' strategies. Next to that, strategies within the trials were coded separately, as participants would sometimes use more than one strategy per trial. Coding strategies within a trial is new and has not been done before. The following categories of strategies were coded: Perceptual motor strategies (Depict all, Depict chain), Abstract strategies (Parity, Incorrect rule), and Other strategies (Guessing, No strategy, Counting, Adding/Subtracting two gears). The different strategies are further explained in Table 2, which is partly adapted from Alibali et al. (2011). The ELAN file with the strategy coding was then saved as .txt.

Table 2

Strategies Used Between and Within the Trials

Strategy used	Definition
Perceptual motor strategies	
Depict all	Participant models the action of each gear.
Depict chain	Participant models the actions of each individual gear.
Abstract strategies	
Parity	Participant notes that if the number of gears is even, the target gear will turn in a direction opposite to the first gear; if the number of gears is odd, the target gear will turn in the same direction as the first.
Incorrect rule	Participant uses an incorrect rule, such as “all gears turn in the same direction”.
Other strategies	
Guessing	Participant states that he or she guessed in order to arrive at a solution.
No strategy	Participant offers the solution only, without stating how the solution was reached.
Counting	Participant counts the number of gears on their fingers without verbally stating the direction of the gear.
Adding/subtracting two gears	The participant mentions the direction of a previously solved gear problem, then adds or subtracts two, such as, “the fourth gear went counterclockwise, so if you then add two, the sixth gear will also go counterclockwise”.

Note. This table contains all strategies that were used for coding and their explanations. It is partly adapted from Alibali et al. (2011)

Next, R studio was used one more time again to extract the recurrence rate of the body movements, in terms of pixel change from frame to frame (Frame Differencing Method). This was done using the crqa-package. Both the embedding dimension and delay were kept to their default of 1. The radius was set to 0.1. The recurrence rates were calculated per participant per trial and added to an excel datafile in which the ELAN .txt file for each participant was also imported We used a custom script in R-studio to combine the audio data, movement data, RQA data, and coded data into timeseries (100 Hz), which were then used to calculate the outcome variables. These outcome variables were combined in one big dataset on which further analyses were done to answer the specific research questions. Lastly, the data were imported into JASP statistical software (Version 0.17.1; JASP team, 2023) for analysis.

Results

The first hypothesis regarded the replication of Alibali and colleagues (2011), in which participants in the gesture-restricted condition were expected to use the parity strategy more often. A Chi-squared test was used to test this hypothesis. The dependent variable for this test was the usage of the parity strategy and the independent variable was the condition that the participant was in (i.e. gesture-allowed or gesture-restricted). In the gesture-allowed condition, 17 participants (50%) used the parity strategy out of 34 participants and in the gesture-restricted condition 16 participants (50%) used the parity strategy out of 32 participants. Inconsistent with the hypothesis, participants in the gesture-restricted condition did not use the parity strategy significantly more often than the gesture-allowed condition, X^2 (1, N = 66) = 0.00, p = 1.00.

The second hypothesis predicted that participants in the gesture-restricted condition had lower recurrence rates in the first trials of the experiment compared to the gesture-allowed condition. This was evaluated using a two-way repeated measures ANOVA, in which the dependent variable was the recurrence rate per trial and the independent variables were condition and parity. Both the assumption of normality for the dependent variable and sphericity were not met; the first was assessed using distribution plots and the latter by Mauchly’s Test of Sphericity, which gave a significant result of $p < .05$. The violation of these assumptions results in an underestimation of the *p*-value of the ANOVA. The results of the repeated measures ANOVA are mentioned in Table 3 and visualized in Figure 2. Figure 2 shows a decline of recurrence rates during the experiment, and these differences in recurrence rates per trial are significant, $p < .01$. However, these differences are similar in both conditions, as we found no significant effect for trials combined with condition, $p = .89$. Moreover, a similar result was found for trials when combined with the use of the parity strategy, $p = .69$. No interaction effect between parity use and condition was found, $p = .75$. Post-hoc tests were conducted after the repeated measures ANOVA to investigate any significant differences between the recurrence rates of both groups per trial. However, all these results were insignificant, $p = 1.00$. The results of the post-hoc comparisons can be found in Table 4.

Table 3

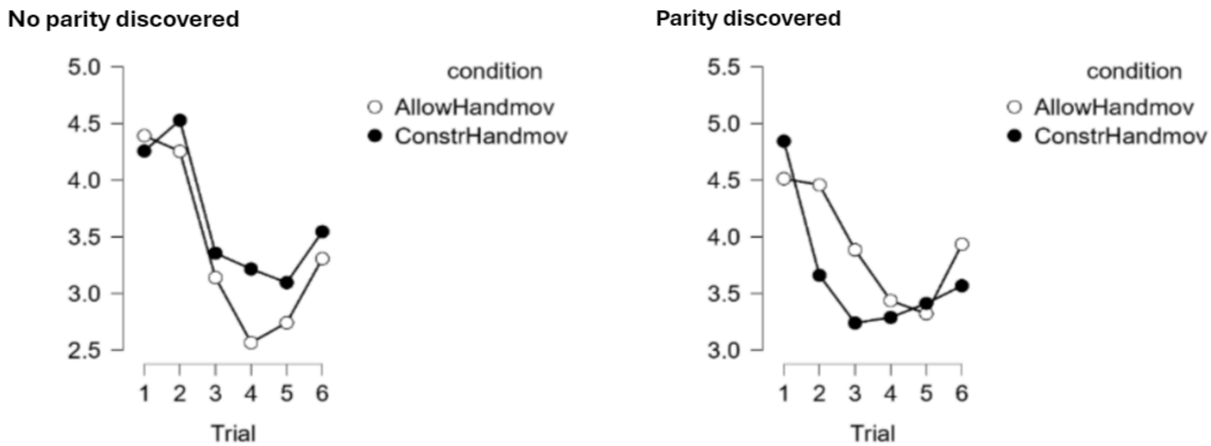
Repeated Measures ANOVA of Trial, Condition and Parity

	Sum of Squares	df	Mean Square	F	p
Trial	83.72	5	16.74	9.06	< .01
Trial * Condition	3.06	5	0.61	0.33	.89
Trial * Parity	5.67	5	1.13	0.61	.69
Trial * Condition * Parity	4.89	5	0.98	0.53	.75
Residuals	434.25	235	1.85		

Note. This table consists of the results of the Repeated Measures ANOVA on hypothesis 2. The sum of squares, degrees of freedom, mean square, F-value and p-value can all be found in the table.

Figure 2

Line Plots of the Average Recurrence Rates per Trial, per Condition



Note. The left figure visualizes the recurrence rates of participants that eventually did not discover the parity strategy and the right figure visualizes the recurrence rates of participants that did discover the parity strategy. In both figures the data of the gesture-allowed condition is shown by the white-dotted line and the gesture-restricted condition by the black-dotted line.

Table 4

Post-Hoc Comparisons of Average Recurrence Rates per Condition, per Trial

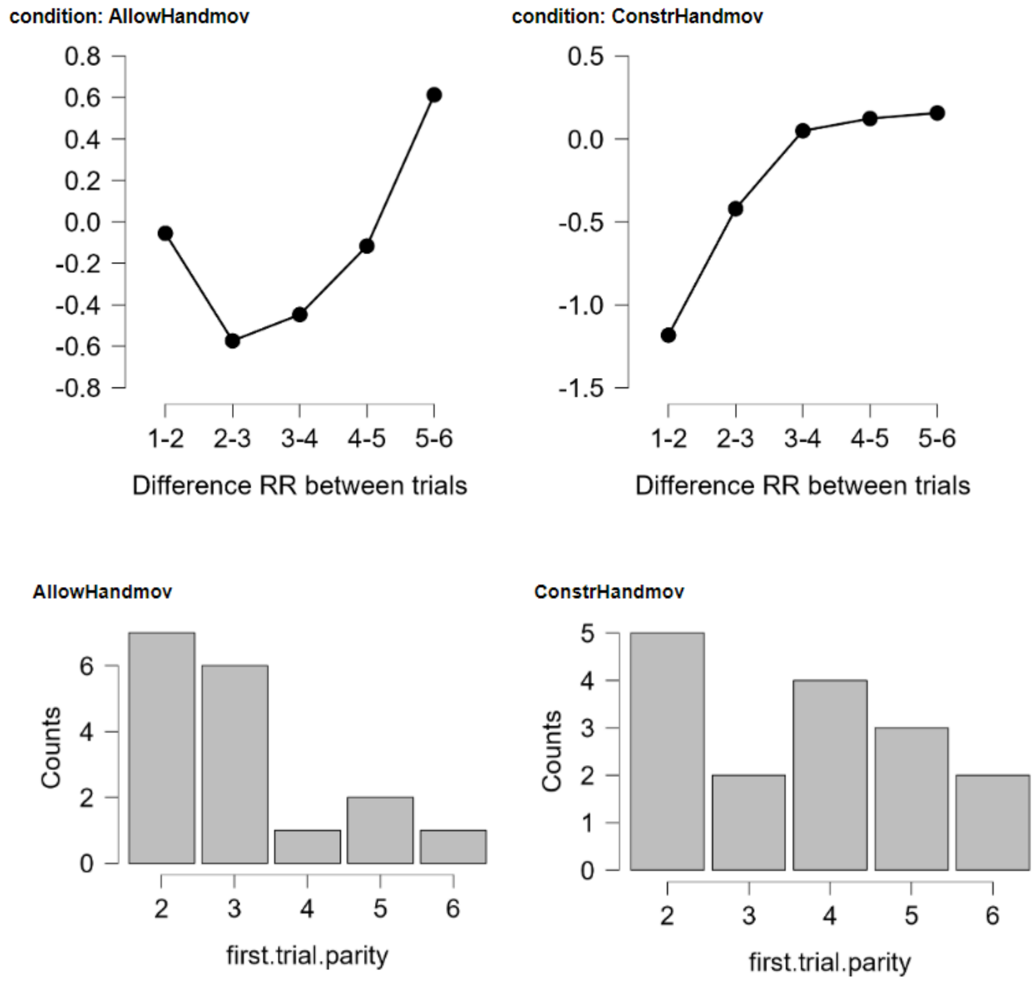
Recurrence Rates – Allowed vs. Restricted Condition	Mean Difference	SE	t	p_{holm}
Trial 1	-0.10	0.46	-0.21	1.00
Trial 2	0.26	0.46	0.57	1.00
Trial 3	0.22	0.46	0.47	1.00
Trial 4	-0.25	0.46	-0.55	1.00
Trial 5	-0.22	0.46	-0.48	1.00
Trial 6	0.06	0.46	0.14	1.00

Note. This table consists of the results on the post hoc tests that were done to compare the recurrence rates of both conditions per trial. The gesture-allowed condition was compared to the gesture-restricted condition. Of these comparisons, the mean difference, standard error, t-value, and p_{holm}-value can be found in the table.

Lastly, an exploratory, visual comparison was done to investigate whether a drop in recurrence rates concurred with the discovery of the parity strategy. This comparison can only be done for the participants who discovered the parity strategy, leaving a total of 24 participants whose data was used (n = 12 for each condition). For this, a total of five new variables were created that expressed the difference in recurrence rates between two trials. For example, the difference between trial 1 and 2 was calculated by subtracting the recurrence rate of trial 1 from the recurrence rate of trial 2, creating the variable diffRR.trial1-2. Variables for trials 2-3, 3-4, 4-5 and 5-6 were created similarly. Two line plots of the mean difference in recurrence rate per two subsequent trials were created, one for each condition. These line plots can be found in the upper part of Figure 3. As a comparison, two histograms, one for each condition, were created that show the number of participants that discovered the parity strategy per trial. These histograms can be found in the lower part of Figure 3, in which each histogram is put beneath the line plot of the same condition. For the gesture-allowed condition, on the left side of Figure 3, a dip in recurrence rates is found before trial 3. Even though one more person discovered the parity strategy in trial 2, there is a high number of participants that discovered the parity strategy in trial 3. For the remainder of the trials, the recurrence rates keep on growing and the parity strategy is discovered less, showing a rather linear relationship between mean recurrence rate differences and the discovery of the parity strategy for this condition. For the gesture-restricted condition, at the right side of Figure 3, a dip in recurrence rates is visualized before the second trial. This corresponds to the high number of participants in that condition that discovered the parity strategy on the second trial. However, for the remainder of the trials recurrence rates do not fluctuate that much anymore and the parity strategy keeps on getting discovered. Therefore, no clear linear relationship between mean recurrence rate differences and the discovery of the parity strategy is found for the gesture-restricted condition.

Figure 3

Gesture-Allowed vs. Gesture-Restricted: Average Recurrence Rates Differences and Parity Strategy Discovery per Trial



Note. The upper part of this figure visualizes the difference in recurrence rates of two subsequent trials for participants that did discover the parity strategy through line plots. The lower part consists of histograms that visualize the distribution of parity-discovery per trial. Not one participant discovered the parity strategy in the first trial, therefore this has been left out. The two figures on the left side correspond to data from the gesture-allowed condition and the two figures on the right side correspond to data from the gesture-restricted condition.

Discussion

Previous research has shown that gestures influence strategy use during the gear-system task (Alibali et al., 2011). The current study consisted of the analysis of two hypotheses together with an explorative visual inspection to investigate the relationship between gesture use and problem solving strategies as well as cognitive variability. First, the study of Alibali et al. (2011) was replicated to check whether the use of the parity strategy differed between the gesture-allowed and gesture-restricted condition. However, the same percentage (50%) of participants used the parity strategy in the gesture-allowed and gesture-restricted condition, and we found no association between condition and parity use. This is not consistent with the hypothesis and conclusions of Alibali et al. (2011), who did find an association between conditions and the use of the parity strategy.

One explanation for this discrepancy between Alibali et al.'s (2011) results and our results could be the many codings of 'no strategy' in the gesture-restricted condition; my colleague Brink (2024) found that in this condition a greater proportion of 'no strategy' codings were present, compared to the gesture-allowed condition for the participants that did not use the parity strategy. In the gesture-restricted condition, it was impossible to interpret the strategy a participant was using by looking at their body movements. When a participant did not explain how they got to their answer, 'no strategy' was coded. It might be possible these participants were using the parity strategy. This, however, is a speculation and future research on the gear-system task should encourage participants more to explain how they came up with their answer. Another explanation for the discrepancy between our and Alibali et al.'s (2011) results is that the power of our study was lower, as our study consisted of a slightly lower number of participants (N = 68) compared to Alibali et al.'s study (2011) (N = 84). Less power gave our study a lower likelihood of rejecting a false null hypothesis, which may have contributed to the insignificant result that we found. However, Alibali et al.'s study and our study have vastly different outcomes that are unlikely to be due to a difference in sample size of 16 participants. Therefore, our replication strongly questions the reliability of the effect found by Alibali et al. (2011). Future research should include more replication studies, to better explore the existence of an effect of gesture-restriction on strategy use during the gear-system task. On other tasks involving gestures, studies have reported a lack of effects similar to ours.

Much research has been done on gestures over the years and more recently studies report no effects of gestures on speech production (Cravotta et al., 2021; Kisa et al., 2022) or problem solving (Cooperrider & Goldin-Meadow, 2014; Walkington et al., 2019). A study in which gestures were restricted, like our study, found that this had no effect on participants' mathematical problem-solving skills (Walkington et al., 2019). Their research supports the *gestures as a byproduct hypothesis* which states that gestures are merely a byproduct of reasoning processes and that gestures therefore do not influence reasoning. Our results are also in line with this hypothesis.

The remainder of the current study derived from a functional dynamic systems' approach. The second hypothesis stated that the gesture-allowed condition was expected to have higher states of variability in the first trials of the gear-system task, compared to the gesture-restricted condition. This was measured through recurrence rates that were obtained through RQA. This hypothesis was not supported by our results. Both conditional groups did not differ on their recurrence rate, regardless of whether they discovered parity. In general, both groups started with high recurrence rates for trial 1 and 2, dropping to lower recurrence rates for trial 3 to 5 and increasing again for trial 6. This trajectory does indicate that variability, measured through recurrence rates, changes during the course of the experiment, which is congruent with the significant result of the variable 'trials' that we found during the repeated measures ANOVA. This indicates that during the gear-system task, fluctuations in variability take place, which is consistent with the notion of Siegler (2007) that variability is inherent to cognitive development.

Research by Stephen, Boncoddo et al. (2009) focussed on entropy values during the gear-system task, which is another measure obtained through RQA (Webber & Zbilut, 2005). One difference between recurrence rates and entropy is that they work in opposite directions. Both an increase in recurrence rates and a decrease of entropy could be interpreted as an increase in variability. Stephen, Boncoddo and colleagues (2009) derived their entropy rates from eye movement behavior. They compared entropy values from participants that did discover the parity strategy to participants that did not discover the parity strategy at all. Stephen, Boncoddo et al. (2009) found a difference in the entropy trajectory during the task between these two groups. The group that discovered the parity strategy experienced peaks in their entropy levels especially before the discovery of the parity strategy, whereas the non-discovery group had a more consistent level of entropy throughout the experiment. This difference between groups in variability is not in line with our findings, as both parity discovery and non-discovery participants did not significantly differ on their recurrence rates throughout the experiment.

One explanation for this incongruence, is the use of the different measures for variability. Entropy is a measure of complexity, indicating the randomness of a system, whereas recurrence rates are based on stability of the system (Webber & Zbilut, 2005). A colleague of mine, Das (2024), investigated entropy rates of our participants. In Das' (202s) study, no significant differences were found in entropy rates between participants that used the parity strategy and those that did not use the parity strategy. This can be seen as an indirect, unsuccessful replication of Stephen, Boncoddo et al. (2009). However, the measures between the two studies are quite different as one uses RQA on eye movements and the other on body movements. The RQA that Stephen, Boncoddo and colleagues (2009) performed on eye movements consisted of longer timeseries, increasing the reliability of the RQA outcomes (Stephen, Dixon & Isenhower, 2009). Moreover, the task itself was different between both studies. Stephen, Boncoddo et al. (2009) challenged participants with a higher number of gear problems which were much more complex than the gear-system task of the current study. Therefore, future research should also focus on replicating these findings, preferably by using a direct replication.

Lastly, a visual inspection compared the difference in recurrence rates on two subsequent trials to the number of participants switching to the parity strategy (Figure 3), to investigate whether switching to the parity strategy was preceded by an increase in variability. As expected in hypothesis 2, the gesture-restricted condition showed the most variability (negative difference in recurrence rates) at the start of the trial, whereas for the gesture-allowed condition this is more distributed over the rest of the trials. Contradictory, the repeated measures ANOVA that we performed regarding this hypothesis indicated no effect of lower recurrence rates for the gesture-restricted condition in earlier trials compared to the gesture-allowed condition. Moreover, we found that as recurrence rates decrease, more parity discovery seems to occur, for both conditions. At later trials recurrence rates increase again and parity is discovered less for the gesture-allowed condition and equally for the gesture-restricted condition. This is merely an explorative inspection of the data, which gives some evidence for hypothesis 2 and a relationship between decreasing recurrence rates and the discovery of the parity strategy (mainly for the gesture-allowed condition). However, it is not a statistical test, so these results should be interpreted with caution. This is especially the case for the differences in variability between conditions at the start of the experiment that were found in Figure 3, as the statistical test that was performed regarding this hypothesis gave no significant result. Future research should further investigate these findings with proper statistical tests.

Our data, apart from the different hypotheses and inspections, had some limitations of its own. All participants were students from the Faculty of Behavioural and Social Sciences from the same university, making our sample less representative of the population and decreasing the generalizability of our results. Next to that, more than half of the participants had to be excluded after participation in the experiment. This is due to the many different complications that arose when watching the videos: Researchers did not follow protocol, cameras were not set up properly et cetera (see Method for a complete overview). However, due to our careful inspection of the videos, it was made sure that participants followed the same treatment, increasing the validity of the results that we found.

In practice, these results, especially the unsuccessful replication of Alibali et al. (2011), lessen the importance that is currently put on the use of gestures during problem solving. There are specific educational programs that encourage gesture use during learning. However, at least for tasks that are similar to the gear-system task, it is necessary to rethink the benefit of actively encouraging gesture use because spontaneous gesture use does not prove to be helpful. For example, a study on spontaneous gesture use during problem solving through analogical reasoning found no positive effect for the use of spontaneous gestures on problem solving (Cooperrider & Goldin-Meadow, 2014). Moreover, spontaneous gestures were found to be detrimental on problem solving during one phase of the experiment. One interesting factor that could contribute to the usefulness of gestures is the instruction of their relevance that is given by the teacher. It has been found that actively encouraging gesture use has beneficial effects on mathematical problem solving only when their relevance is mentioned by the instructor (Brooks & Goldin-Meadow, 2016). This could be a boundary condition for the usefulness of gestures. Therefore, future research should focus on the careful replication of studies that found a positive effect for gesture use in learning or problem solving as well as exploring the boundary conditions that allow gestures to be helpful.

Furthermore, future research should apply a functional dynamic point of view to gain more insight in the cognitive system. Specifically, research could focus on further exploring how strategy switches and variability interact. For more reliable RQA outcomes, the research paradigm of Stephen, Boncoddo et al. (2009) could be replicated with the extraction of recurrence rates as well as entropy from the RQA. Variability during the different trials of the gear-system task could be tested to gain more insight in the specific trajectory of stability or randomness of the cognitive system, especially when strategy switching takes place. Moreover, stability and randomness of the system could be compared through the comparison of recurrence rates and entropy, giving more insight into the differences between these two measures to decrease the gap between research that uses either of these.

References

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous Gestures Influence Strategy Choices in Problem Solving. *Psychological Science*, *22*(9), 1138-1144. https://doi.org/10.1177/0956797611417722

Arzarello, F., Paola, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Mathematics*, *70*, 97-109. https://doi.org/10.1007/s10649-008-9163-z

Brink, M. (2024) *Invloed van de hoeveelheid gebaren op het probleemoplossend vermogen* [Bachelor's thesis, University of Groningen]

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

Bross, F. (2020). Why do we shake our heads? On the origin of the headshake. *Gesture, 19*(2-3), 269-298. https://doi.org/10.1075/gest.17001.bro

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*(1), 102–116. https://doi.org/10.1037/a0021790

Cooperrider, K., & Goldin-Meadow, S. (2014). The role of gesture in analogical problem solving. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 36*(36), 2068-2072.

Cravotta, A., Prieto, P., & Busà, M. G. (2021). Exploring the effects of restraining the use of gestures on narrative speech. *Speech Communication, 135*, 25-36. https://doi.org/10.1016/j.specom.2021.09.005

Das, S. (2024) *Temporal Patterns in Attractor States to Understand Problem-Solving* [Bachelor's thesis, University of Groningen]

De Bari, B., & Dixon, J. (2022). Dissipative structures as an alternative to the machine metaphor of the mind and brain.

de Jonge-Hoekstra, L. (2021). *How hand movements and speech tip the balance in cognitive development: A story about children, complexity, coordination, and affordances*. [Thesis fully internal (DIV), Rijksuniversiteit Groningen]. University of Groningen. https://doi.org/10.33612/diss.172252039

JASP Team (2023). JASP (Version 0.17.1)[Computer software].

Kısa, Y. D., Goldin-Meadow, S., & Casasanto, D. (2022). Do gestures really facilitate speech production? *Journal of Experimental Psychology: General, 151*(6), 1252–1271. https://doi.org/10.1037/xge0001135

Kondepudi, D. K., De Bari, B., & Dixon, J. A. (2020). Dissipative structures, organisms and evolution. *Entropy*, *22*(11), 1305.

McNeill, D. (2005). *Gesture and thought.* University of Chicago Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Paxton, A., & Dale, R. (2012). Frame-differencing methods for measuring bodily synchrony in conversation. Behavior Research Methods, 45(2), 329–343. https://doi.org/10.3758/s13428-012-0249-2

Pouw, W., Trujillo, J. P. (2019). Tutorial Gespin2019 - Using video-based motion tracking to quantify speech-gesture synchrony. https://doi.org/10.17605/OSF.IO/RXB8J

Schrödinger, E. (1944). *What is Life?* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139644129

Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current directions in psychological science, 3*(1), 1-5. https://doi.org/10.1111/1467-8721.ep10769817

Stephen, D. G., Boncoddo, 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. (2009). The self-organization of insight: Entropy and power laws in problem solving. *Journal of Problem Solving*, *2*(1), 72-102. https://doi.org/10.7771/1932-6246.1043

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1811-1832. https://doi.org/10.1037/a0014510

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

Walkington, C., Woods, D., Nathan, M. J., Chelule, G., & Wang, M. (2019). Does restricting hand gestures impair mathematical reasoning?. *Learning and Instruction, 64,* 101225. https://doi.org/10.1016/j.learninstruc.2019.101225

Webber Jr, C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, *94*(2005), 26-94.

Will, I. (2021). Chapter 1 Function, Meaning and Form of Gestures. *Recurrent Gestures of Hausa Speakers*. (pp. 7-31) Brill. https://doi.org/10.1163/9789004449794_003

Zurina, H., & Williams, J. (2011). Gesturing for oneself. *Educational Studies in Mathematics*, *77*, 175-188. https://doi.org/10.1007/s10649-010-9294-x