Towards Extraction of Cognitive Stages of Processing During Mental Spatial Processing With Gaze Data of a Mental Folding Task

Anonymous Author(s)

Figure 1: A screenshot of a mental folding trial.

ABSTRACT

Distinct cognitive stages of processing during mental spatial transformation tasks may be evident in oculomotor behavior. We recorded eye movements whilst participants performed an altered version of the mental folding task. Gaze behaviour was analyzed to provide insight on the relationship between task difficulty, gaze proportion on each stimuli, gaze switches between stimuli and reaction times. We found a linear decrease in switch frequency and gaze proportions with increasing difficulty level. Further, we demonstrate that these measures of gaze behaviour are related to the time taken to perform the mental transformation. We propose that the observed patterns of eye movements are indicative of distinct cognitive stages during mental folding. Lastly, further analyses with exploratory methods as well as possibilities for cognitive modeling are discussed.

CCS CONCEPTS

• **Computing methodologies** → *Cognitive science*; **Spatial and physical reasoning**; • **Applied computing** → *Psychology*.

KEYWORDS

mental spatial transformation, mental folding, cognitive stages of processing

ACM Reference Format:

Anonymous Author(s). 2020. Towards Extraction of Cognitive Stages of Processing During Mental Spatial Processing With Gaze Data of a Mental Folding Task. In *Proceedings of The 12th ACM Symposium on Eye Tracking Research and Applications (ETRA 2020).* ACM, New York, NY, USA, 5 pages. https://doi.org/xx.xxxx/xxxxxxx.xxxxxxx

ETRA 2020, June 2–5 2020, Stuttgart, Germany

ACM ISBN XXX-X-XXXX-XXXX-X/XX/XX…$XX.XX
https://doi.org/xx.xxxx/xxxxxxx.xxxxxxx

1 INTRODUCTION

The ability to imagine physical interaction with objects in space and to assess, plan and execute actions based on this mental representation is a fundamental aspect of human life. This ability emerges from the interplay of multiple cognitive processes. These processes of mental spatial transformation are distinctly separate from non-transformative processes such as navigation or orientation [Harris et al. 2013]: the object of investigation is transformed in a way that potentially changes its behavior, function and shape.

One subfield of cognitive processes involved in mental spatial transformations is mental folding. A typical paradigm to investigate mental folding, originally by Shepard and Feng [1972], requires participants to compare the surfaces of a cube - the "reference" stimulus - to a folding pattern shown in parallel - the "target" stimulus - and subsequently decide if specific patterns on the cube can be a result of the folding pattern if folded up. Difficulty of the task can be varied by increasing the number of folding processes necessary to reach a conclusion. This instruction forces a test participant to simulate the folding process mentally instead of physically and helps to analyze mental spatial transformation in isolation. While similar to mental rotation experiments [Shepard and Metzler 1971], mental folding has a distinctive property in facilitating change of single attributes or aspects of an object, rather than simply manipulating an object as a whole.

An understanding of mental spatial transformation has great potential especially for economic applications, such as improvement of product ergonomics (e.g. improving ease of use and overtness of features), or accessibility (e.g. identifying (dis-)advantages of individual traits). To this end, it is important to gain insight into cognitive stages of processing that are undergone during mental object interaction. Knowledge of cognitive stages enables a more guided approach on understanding human-machine interaction. Additionally, an informed process model of spatial transformative

cognition can serve as an outline for the development of cognitive models, using a cognitive architecture such as ACT-R [Anderson et al. 2004]. These models can then, in turn, provide predictions and assessments of mental spatial transformation processes and extract inter-individual differences and common features in cognition. The investigation of mental folding can resolve much of the uncertainty about these processes, as object transformation is crucial for mental spatial cognition.

Noton and Stark [1971] showed that processes of visual encoding and learning are reflected in eye movements. Just and Carpenter [1976] demonstrated further that separate stages of cognitive processing are distinguished by differential patterns of fixations. In a mental rotation paradigm similar to Shepard and Metzler [1971], they observed that the number of fixation switches between the two presented figures was increasing linearly with the angle offset of the figures, which served as an analog to processing difficulty: specifically, gazes on single aspects of presented stimuli seem to coincide with encoding processes of these aspects, with more complex figures taking more time to process. Based on this and the scan paths over stimuli features, they formulated three distinct stages of processing during the mental rotation task:

- the **initial search process** to find corresponding features between figures,
- the **transformation and comparison process** to align and compare the two figures and
- the **confirmation process** to confirm sameness of the presented stimuli.

These three steps are generalizable and could be considered a basic processing model for mental spatial transformation [Just and Carpenter 1976]. Kosslyn [1996] proposed activation mechanisms for mental representations that are similar to those found in declarative memory models: renewed encoding of the origin of a mental representation helps to actively maintain it. This is in line with results by Cowan [1999] who found mechanisms for active maintenance in working memory. It follows that if stimulus fixations are understood as correlates of encoding, comparison tasks with higher difficulty should show a decrease in the proportion of fixations on reference stimuli, as encoding/initial search processes decrease in importance while transformation processes, which are focussed on target stimuli, increase. Other researchers elaborated on the relation between object complexity and fixation time, for example when reading, fixation times increase when reading words of greater lexical complexity on fixation times [Rayner and Duffy 1986].

These approaches on cognitive processing stage extraction have not been applied to a mental folding paradigm so far. In this paper, we present analyses of gaze data gathered during a mental folding experiment as indications for time- and condition-specific cognitive processing stages. We expect our analyses to show **1)** differences in the amount of gaze switches across difficulty levels and **2)** differences in the proportion of time spent looking at the reference stimulus across difficulty levels [Just and Carpenter 1976][Kosslyn 1996][Rayner and Duffy 1986].

This should serve as a proof-of-concept for extracting a relation between eye movements and task processing. Following up on this, we plan to apply more complex modeling approaches to our data:

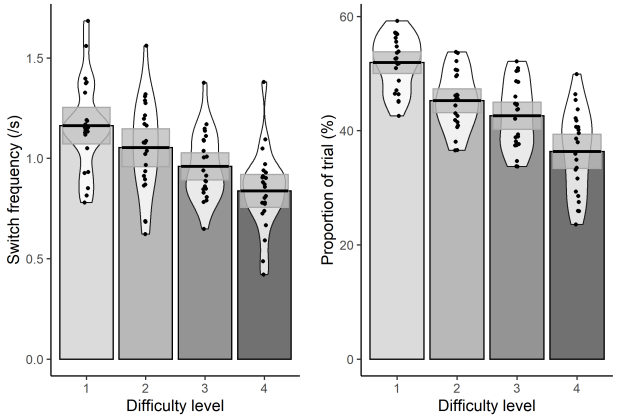

Figure 2: *Left:* Effect of difficulty level on gaze switch frequency. *Right:* Effect of difficulty level on reference stimulus gaze proportion.

several exploratory approaches for the extraction of cognitive stages from eye tracking data have been proposed in recent years, for instance hidden Markov models [Xue et al. 2017][Chan et al. 2019], pupil dilation analysis [Wierda et al. 2012] or Gaussian mixture models [Feng 2006]. Their feasibility for use on our current and future data is discussed in the conclusion of this paper.

2 METHODS

23 participants (10 female) took part in eye tracking. Average age of participants was 29.04 (SD = 5.17).

2.1 Experiment Design

A computerized version of the mental folding task originally developed by Shepard and Feng [1972] in a variant by Wright et al. [2008] was presented to the participants. The task was presented in the form of semitransparent 3D cubes as reference figures and 2D unfolded cube templates as target figures presented on a black background, each with two black arrows on their surfaces and a blue square indicating the base. Each trial started with a one second presentation of a central fixation cross, followed by the display of the reference figure, either on the left or right side of the screen. After an additional second, the target figure appeared on the other respective screen side. Participants were asked to mentally fold the template into a cube shape and to decide whether the arrows on reference and target match. Judgements on matching or mismatching arrow positions were recorded via button presses on a response pad, with vertically aligned buttons for match and mismatch answers. The experiment consisted of 600 trials, subdivided in five blocks. Participants had to take at least one minute breaks between the blocks and were instructed to always fold upwards, starting from the base. Task completion took 60 minutes on average. In advance, each participant passed through 10 minutes of training with feedback on correctness of their answer.

Trials are divided into four levels of difficulty. The number of squares carried during the series of folds necessary to compare the

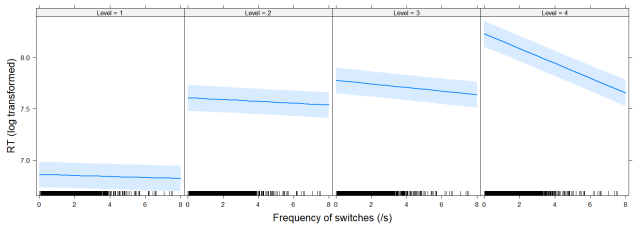

Figure 3: LME predictions of log transformed reaction times as function of switch frequency (/s) between the reference and target objects across different difficulty levels. Line ribbons show 95% confidence intervals.

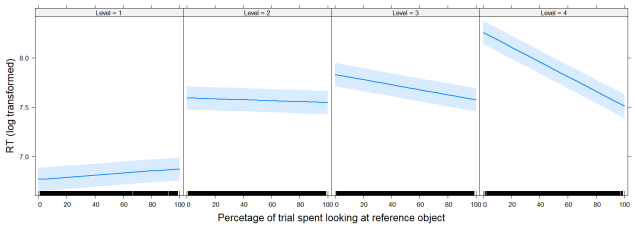

Figure 4: LME predictions of log transformed reaction times as function of gaze proportion on reference stimulus across different difficulty levels. Line ribbons show 95% confidence intervals.

arrow positions determined the level of difficulty. The easiest level was a direct visual comparison with arrow tips always meeting and one of the arrows located on the base square. The second level required carrying 4, the third level 5 and the fourth level 6 squares through the folding sequence. Six different pattern templates with three arrow variations each (for Levels 2, 3 and 4: one variation with arrow tips touching, two with arrow tips in different directions) were constructed for every difficulty and paired with reference cube figures with either matching or mismatching arrow positions. This resulted in 144 different trials. To shorten the length of the experiment to one hour, 24 trials of the mismatch condition were excluded by balanced randomization from each block. Each mismatch stimulus type of each level was shown at least three times over the whole experiment, resulting in 72 match- and 48 mismatch trials per block. The sequence of trials and the presentation sides were randomized in a balanced manner within each block. The design of the first difficulty level is of particular note for the differentiation of cognitive stages: as it can be solved by direct visual comparison, it bypasses the need for a distinct stage for mental spatial transformative processing, thereby facilitating more straightforward solving strategies that might be reflected in gaze patterns.

2.2 Eye tracking

Eye tracking data was collected through a *The Eye Tribe* tracker, at a sampling rate of 30Hz. Sampling rate was too low to parse gaze data into fixation, saccades and blinks. As a result, we analysed raw gaze data. Out of 13800 total trials, 696 trials were removed due to missing gaze data. Our analysis focused on whether gaze was directed towards the reference or the target object, for which we used the following measures:

- **Gaze switches**: switches between reference or target object were counted when gaze was on another side of the threshold than the preceding gaze point.
- **Gaze duration**: gaze duration refers to the proportion of time spent looking at the stimulus. Only reference gaze duration is reported, as target gaze duration is its exact inverse.

3 RESULTS

We analyzed the data using Linear Mixed Effects models (LME) in R [R Core Team 2019] using the LME4 package (version 1.1-21 [Bates et al. 2015]). For each model we began with a maximal random effects structure and iteratively reduced by-participant slopes for fixed effects in order to reduce model complexity (see Bates et al. [2018]). The final model was selected based on AIC comparison between models. Using this procedure, the intercept only models were chosen in all cases. t-values at 1.96 or higher were considered as statistically significant.

3.1 Gaze Switches

Absolute number of gaze switches increased linearly with difficulty level (mean switches level one: 2.45, two: 3.45; three: 3.58; four: 3,96). However this was confounded by trial length (which increased linearly with difficulty). Instead, we calculated switch frequency, which was the number of switches per second. We conducted an LME on switch frequency with difficulty level as fixed effect (factor; level 1,2,3,4; coded using successive difference contrasts) and participant as a random effect. The model revealed a significant linear decline in switch frequency with increasing difficulty (see Figure 2 *(left)*) from level 1 to 2 ($\beta = -0.11, SE = 0.02, t = -5.61$), 2 to 3 ($\beta = -0.09, SE = 0.2, t = -5.61$) and 3 to 4 ($\beta = -0.12, SE = 0.02, t = -7.12$).

To investigate whether this change in switching behaviour across difficulty level was related to task performance, we performed a follow up analysis to analyse whether gaze switching was a predictor of reaction times. We conducted an LME on reaction times (log-transformed), with difficulty level (factor; level 1,2,3,4; coded using successive difference contrasts) and switch frequency (continuous, centred) as fixed effects and participant as a random effect. Firstly, the model showed that reaction times increased linearly with task difficulty from level 1 to 2 ($\beta = 0.73, SE = 0.01, t = 57.55$), 2 to 3 ($\beta = 0.14, SE = 0.01, t = 10.67$) and 3 to 4 ($\beta = 0.24, SE = 0.01, t = 18.31$).

Switch frequency was a significant predictor of reaction times ($\beta = -0.08, SE = 0.01, t = -14.50$) such that a higher switch frequency corresponded with shorter reaction times. However this effect was qualified by an interaction with difficulty level (see Figure 3). Specifically, the relationship between switch frequency and reaction time was almost non existent at the easier difficulty levels and did not change from level 1 to 2 ($\beta = -0.01, SE = 0.01, t = -0.94$) and 2 to 3 ($\beta = -0.03, SE = 0.02, t = -1.68$). However, the relationship is much stronger by level 4 and this increase from level 3 was significant ($\beta = -0.16, SE = 0.02, t = -9.96$).

3.2 Gaze Duration

For gaze duration on objects, we present analysis for gaze duration on reference object only, since gaze on target object is the inverse in this measure. Specifically, gaze duration was conceptualised at the proportion of trial spent looking at the reference object, in order to account for varying trial lengths. We conducted an LME on gaze duration with difficulty level as a fixed effect (factor; level 1,2,3,4; coded using successive difference contrasts) and participant as a random effect. The model revealed a significant linear decline in gaze duration with increasing difficulty (see Figure 2 *(right)*) from level 1 to 2 ($\beta = -6.60, SE = 0.47, t = -14.08$), 2 to 3 ($\beta = -2.63, SE = 0.47, t = -5.62$) and 3 to 4 ($\beta = -6.17, SE = 0.47, t = -13.18$).

Gaze duration on the reference object, as a proportion of trial time, was a significant predictor of reaction times ($\beta = -0.08, SE = 0.01, t = -14.50$) such that larger proportions of the trial spent looking at the reference object related to shorted reaction times. However, this effect was qualified by an interaction with difficulty level (see Figure 4). Specifically, the relationship between gaze duration and reaction time was weak in the easier difficulty levels, and significantly increased in magnitude with advancing difficulty level from 1 to 2 ($\beta = -0.06, SE = 0.01, t = -4.73$), 2 to 3 ($\beta = -0.08, SE = 0.01, t = -5.77$) and 3 to 4 ($\beta = -0.19, SE = 0.01, t = -13.49$).

4 DISCUSSION

To discover evidence for cognitive stages of processing, our hypotheses required us to find different gaze switch behavior as well as different proportion of stimuli gazes across difficulty levels, for both of which we found significant effects.

Gaze switch frequency showed a linear decline with increasing difficulty level, which suggests different solving strategies: almost equal time is spent looking at reference and target stimuli in the easiest condition, which is accompanied by a high switch rate. This suggests that participants are performing a simple perceptual comparison between the two objects. As difficulty increases and (more complex) mental transformations have to be performed, participants shift their attention to the target. More demanding transformation processes seem to require participants to look at the target more to maintain their mental representation, in line with the concept of active maintenance [Kosslyn 1996]. Higher difficulties induce "locking" on the target stimulus to facilitate transformation, explaining the decreasing switch frequency.

Higher absolute numbers of switches occurring with longer reaction time seems to be an indication of a prolonged planning or *initial search* phase. Trials requiring complex transformations take more time to assess. While this interpretation fits the presented data, assumptions about the time course of trials are made - switches at the beginning for planning, target fixation during transformation, switches at the end for comparison - for which evidence is not yet available. For this, methods illuminating gaze behavior over time are required which would give additional insights into the nature of cognitive processes involved with mental spatial transformation.

Several researchers used hidden Markov chains as a model of eye movement patterns (Hidden Markov models, *HMMs*). Chan et al. [2019] established a method for the analysis of scan path similarity between participants. With this quantitative measure, ROIs and their order of fixation can be extracted post-hoc. A similar approach was chosen by Xue et al. [2017], with an extension called discriminative HMMs, for cognitive stage extraction. Their study design however had little constraints regarding differentiation of the stimuli - contrary to our study, there was no pause of 1 second after reference presentation. This might interfere with the applicability of this method, as our study design enforces distinct encoding and transformation stages. Additionally, HMMs assume that fixation sequences are independent to prior sequences [Salvucci 1999], which might be incompatible with our assumption of subsequent cognitive stages.

Another possible application for HMMs is change in pupil dilation. Wierda et al. [2012] connect pupil size increases to mental effort. By assuming specific events in time at which mental effort could be triggered, they were able to generate a fit to pupil responses. Effects of mental processes on pupil dilation are slow however, which could prove this approach infeasible for the experiment design presented in this paper. The method is further shown to be useful for tasks with several predetermined events (such as the AB task used by Wierda et al. [2012]), while our study design can only be separated into events for reference and target stimulus onset.

A promising approach lies in the use of Gaussian mixture models (*GMMs*). Here, a certain data distribution is sought to be fit with multiple Gaussian distributions [Feng 2006]. For the extraction of cognitive processing during mental spatial transformation tasks, the resulting mixture of distributions can then be interpreted as signifying stages of such processing. Furthermore, different strategies of mental folding might exhibit differing patterns of distributions and strengthen the assumption of strategy differences between difficulty levels, e.g. a predominantly visual approach for the first difficulty level and spatial transformative strategies for higher difficulties. Feng [2006] showed the feasibility of GMMs on fixation duration. A possible extension to this would use event-related gaze switch times, so as to be comparable for other measures of cognitive stages extraction. This would use gaze switches respective to the onset of the target stimulus for each participant to show increases or decreases of switching during certain times of the trials. GMMs for subgroups of the data, e.g. separated by difficulty level, could then be juxtaposed by the number and placement of Gaussian distributions to illuminate meaningful differences. So far only used on reading fixations in the eye tracking domain [Feng 2006], an application on event-related switches must be considered exploratory for now.

An upcoming study will aim to combine mental rotation and folding into a single task design. With this, a more detailed look into mental spatial transformation should be possible. For this experiment, an eye tracking setup with a higher sampling rate will be used. This will enable access to parameters not available with the current data, such as fixations, saccades or blink detection. Data gathered from these studies will furthermore aid in the development of cognitive models on mental spatial transformation. By informing improved process models and pinpointing common processes of mental spatial transformation, eye tracking data supports the simulation and prediction of this ability and may ultimately support a unified methodological framework for spatial transformative cognition.

REFERENCES

John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. 2004. An Integrated Theory of the Mind. *Psychological Review* 111, 4 (2004), 1036–1060.

Douglas Bates, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2018. Parsimonious Mixed Models. *arXiv:1506.04967 [stat]* (May 2018).

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

F. Chan, T. Barry, Antoni Chan, and Janet Hsiao. 2019. Hidden Markov modelling of eye movements in social anxiety: a data-driven machine-learning approach to eye-tracking research in psychopathology. In *2019 ADAA Annual Conference*.

Nelson Cowan. 1999. An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control* 20 (1999), 506.

Gary Feng. 2006. Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research* 7, 1 (March 2006), 70–95. https://doi.org/10.1016/j.cogsys.2005.07.004

Justin Harris, Kathy Hirsh-Pasek, and Nora S. Newcombe. 2013. Understanding spatial transformations: similarities and differences between mental rotation and mental folding. *Cognitive Processing* 14, 2 (May 2013), 105–115. https://doi.org/10.1007/s10339-013-0544-6

Marcel A. Just and Patricia A. Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 4 (Oct. 1976), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Stephen M. Kosslyn. 1996. *Image and brain: The resolution of the imagery debate: The resolution of the imagery debate* (1st ed ed.). MIT Press, Cambridge, MA.

David Noton and Lawrence Stark. 1971. Eye movements and visual perception. *Scientific American* 224, 6 (1971), 34–43.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Keith Rayner and Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14, 3 (May 1986), 191–201. https://doi.org/10.3758/BF03197692

Dario D Salvucci. 1999. *Mapping eye movements to cognitive processes*. Ph.D. Dissertation. Carnegie Mellon University, Pittsburgh, PA, USA. Advisor(s) Anderson, John R.

Roger N. Shepard and Christine Feng. 1972. A chronometric study of mental paper folding. *Cognitive Psychology* 3, 2 (1972), 228–243.

Roger N. Shepard and Jacqueline Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Science* 171 (1971), 701–703.

Stefan M. Wierda, Hedderik van Rijn, Niels A. Taatgen, and Sander Martens. 2012. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences* 109, 22 (May 2012), 8456–8460. https://doi.org/10.1073/pnas.1201858109

Rebecca Wright, William L Thompson, Giorgio Ganis, Nora S Newcombe, and Stephen M Kosslyn. 2008. Training generalized spatial skills. *Psychonomic Bulletin & Review* 15, 4 (2008), 763–771.

Jiguo Xue, Chunyong Li, Cheng Quan, Yiming Lu, Jingwei Yue, and Chenggang Zhang. 2017. Uncovering the cognitive processes underlying mental rotation: an eye-movement study. *Scientific Reports* 7, 1 (Dec. 2017). https://doi.org/10.1038/s41598-017-10683-6