Manual Movement during Categorization of Atypical exemplars

A replication study of ‘Graded motor responses in the time course of categorizing atypical exemplars’ (Dale et al, 2007)

Manar Ali, Leon Heidkamp, Marcel Klehr, Berit Reise

Experimental Psychology Lab, Summer Term 2020, University of Osnabrück

31st of July, 2020

Abstract

Categorizing entities is a common cognitive process and daily practice for human beings. Mostly, categorization seems to be fast and intuitive, but objects that represent atypical exemplars of their category can emerge to be more challenging due to multiple categories competing in the categorization process. Dale, Kehoe, & Spivey (2007) investigated the categorization of atypical exemplars with focus on the time course of the process. Here, we replicated the first experiment of their study as an online experiment: Participants were presented with animal names and had to decide between two categories by clicking on the category name that the animal belongs to. Looking at the trajectories of the mouse movement, we report that there is a significant difference between the trajectory curvatures for typical and atypical animals. Our results confirm the ones from the original study.

Keywords: categorization, typicality, mouse-tracking

Introduction

Knowledge representation is an important aspect of human cognition. In order to generalize knowledge and make it universally understandable, comparable and learnable, we use concepts and categories. A concept is a representation of an entities' abstract meaning. It provides us with a general idea about its properties such that we can identify whether an instance fits a concept or not. A category is the collection of all instances that fit a certain concept. Investigating how we acquire such conceptual knowledge and use it to guide categorization in our daily lives has been a major topic in research.

One interesting issue that emerged from this field of research is that not all instances have a unique membership, i.e. belong to only one category. Some instances do not match the underlying concept of their category perfectly, hence they are not a typical example of their category, and they share conceptual features with other categories, meaning that multiple categories compete with each other during the categorization process. These instances exhibit more of a fuzzy membership for their category and seem to be harder to categorize, which leads to the question of how exactly we categorize an instance according to our conceptual knowledge, and how we deal with the problem of (a)typicality during categorization. Research trying to answer these questions diverged considerably and gave rise to the development of different theories of categorization (for an overview see Goldstone, Kersten, & Carvalho, 2012: pp. 611-619).

In addition to multiple theoretical approaches, experimental methods developed further as well. Initial research focused more on the outcome of categorization than the whole process of it. (Dale, Kehoe, & Spivey, 2007: p.15 f.) More recent studies, however, started to investigate the temporal dynamics of categorization processes and take physical movements (e.g. eye movement, hand movement) as indicators of the underlying cognitive processes into account. This line of research exploits the notion that our movements are continuously guided and updated by underlying cognitive processes, i.e. observing those movements can tell us something about the responsible cognitive processes as well (see Freeman, Dale, & Farmer, 2011 for hand movement studies in particular).

Dale, Kehoe, & Spivey (2007) investigated the temporal dynamics of decision processes in a mouse tracking study. In the first experiment of their study, they looked at the categorization process of animal names with different ratios of typicality relative to their category. They compared atypical exemplars (e.g. penguin) that exhibit competing categories (are penguins mammals *or* birds?) with typical exemplars (e.g. dog) that show an unambiguous categorization (mammal).

In this study we present a replication of the first experiment done by Dale, Kehoe, & Spivey (2007). We examined the categorization processes of typical vs atypical animals and the time course of the categorization process by keeping track of the hand, i.e. mouse movement, while participants choose a category for a presented exemplar. The underlying intuition was that the uncertainty in categorizing atypical exemplars due to the competition of categories should be evident in the mouse movement trajectories. For example, when categorizing the animal *penguin*, one's mouse movement might drift to the category *mammal* first, because penguins do not fly like birds usually do, i.e. penguins do not fit the concept of birds perfectly which hampers the categorization process. But after some more contemplation, one will categorize them as *bird* and the mouse will change its path to the right category. Compared to the categorization of *dog*, where we are immediately certain that it belongs to *mammal*, the mouse movement will be quite straight with less divergence. Thus, the mental uncertainty in categorizing an atypical exemplar should also be visible physically, in the staggering hand movements of the participants.

Here, we tested whether there is a significant difference in the categorization process between those two conditions by hypothesising that mouse trajectories should gravitate more towards the competing category (incorrect category) in experimental trials (atypical condition) than in control trials (typical condition).

Methods

Participants

Students of the University of Osnabrueck and people of their close social environment were recruited via email and text messenger by sending them the link to the browser-based study. The participants did not receive compensation for completing the experiment.

Materials

In the experiment each participant was tested on two within-subject conditions: A control and an experimental condition. This means all of the experiment's trials contained the exact same stimuli for all participants. We tried to follow the original experimental design as closely as possible, therefore we adopted the material provided by Dale, Kehoe, & Spivey (2007) for the target trials that were presented to participants in the original experiment. Each target trial staged the name of an animal that needed to be categorized and two possible categories to choose from. The 19 target trials comprised 6 atypical animals (experimental trials) and 13 highly typical ones (control trials) - the animals were typical or atypical in regards to a certain category. There were five different categories: Mammal, reptile, bird, fish, insect and amphibia. In the experimental trials the atypical animal was more difficult to categorize since both provided categories shared conceptual features with the animal. Additionally, we created new stimuli for practice trials.

A central difference is that our study was implemented as an online experiment rather than as laboratory study. We tried to diminish divergences by means of our extended exclusion criteria as described in the section *Data Analysis* in order to replicate the laboratory conditions as closely as possible. Furthermore, we provided precise and thorough written instructions to prevent other difficulties, which are described in the following part.

Procedure

Each participant completed three practice trials before the 19 target trials. For each trial, the two possible categories for the animal name were randomly assigned to one of the upper corners of the screen. Participants saw the category names in their assigned positions for 2000 msec. After that, a text field labeled ‘Start’ appeared at the bottom center which elicited the appearance of the animal name as soon as participants clicked it. When the animal name was shown in the bottom center of the screen, participants had to decide which category it belonged to and select the appropriate one with a mouse click in the corresponding corner. Once a choice was made, a new trial started.

After practice, 19 target trials followed in the same manner. All stimuli were presented in random order and neither in the practice nor in the main trials the participants were provided feedback about whether their category choice was correct or incorrect.

After the main experiment ended, participants had to answer a questionnaire that corresponded to the content of our exclusion criteria such that we were able to rule out participants who did not fulfill the requirements to participate.

Data Analysis

Following the statistical analysis described in our pre-registered report (can be found here), we used R packages tidyverse and zoo to analyse the collected data. We excluded participants that either did not use a computer mouse; or reported that they used their left hand during the experiment (we do not ask specifically about left-handedness but about which hand they used; this exclusion criteria was added to remain as close to the original study as possible where all participants were using their right hand due to the laboratory setup); or participants that aborted the experiment; or who reported a score of two or less (out of seven) in the english comprehension question in the post-experiment questionnaire. Additionally, we excluded data from individual trials when they responded incorrectly or exceeded the time limit of 5.0 seconds.

We recorded x- and y- coordinates from the mouse movement as our dependent variable and typicality (typical vs atypical) as independent variable. Moreover, we measured reaction times and correctness of responses.

With the help of those values, we tested the following hypotheses: For the time-normalized analysis, we were interested in whether the difference between the time-normalized typical and atypical condition's x-coordinates is or is not significant ($p<.05$) in at least 8 consecutive time steps (of 101). For the space-normalized analysis, we wanted to proof whether there is or is not a significant effect ($p<.05$) of trial type on the variance of mouse coordinates.

Results

A total of 68 persons participated of which 29 participants (14 women, 11 men, 4 other) were eligible for the analysis based on our exclusion criteria as described previously. This deviated from our aim of getting 41 eligible participants but due to time limitations, we were not able to recruit more people.

The standard $p<.05$ criteria was used in the following tests.
First, the average accuracy in percentage of the typical and atypical conditions was calculated and compared between the two conditions. The conducted Pearson's χ^2 test with

Yates' continuity correction revealed a significant difference between responses of the two conditions [χ^2 = 14.589, df = 1, p-value = 0.0001337].

To facilitate further analysis, all leftward responses were reflected at the y-axis. For the time-normalized part, the data was normalized to 101 time steps and all start points to the point (0,0). For all 101 time steps, x-coordinates were averaged per condition and then compared by applying a t-test. Results show that there are significant differences between conditions in more than 8 consecutive time steps, hence we accept H1 of our time-normalized hypothesis.

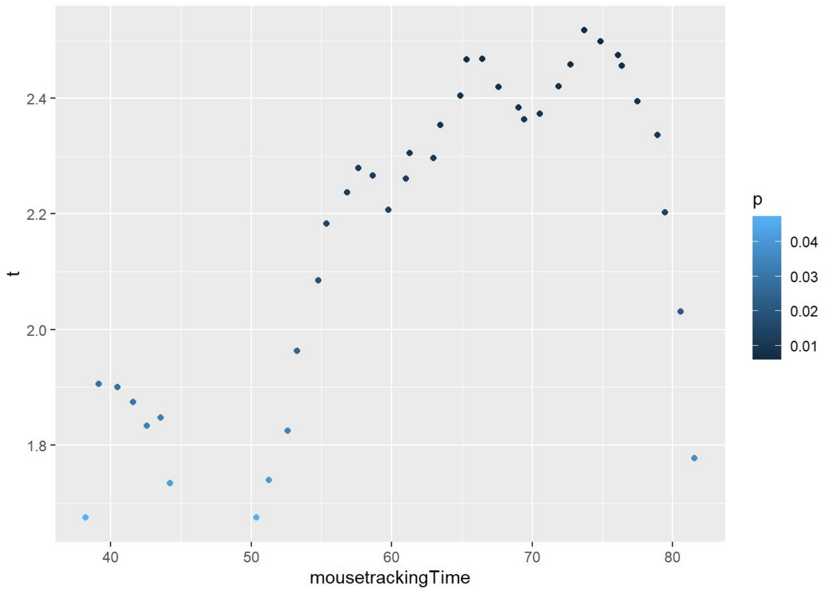

Figure 1. Significant t-test Results between Conditions (t- values on y-axis; 101 time steps on x-axis)

Points indicate statistically significant differences between conditions for a time step; significant p-value is indicated by tint of blue. Note, that there are more than 8 points depicted, hence more than 8 consecutive time steps revealed a significant difference.

In addition to that, we separated the 101 time steps into three bins (1-33; 24-67; 68-101) and applied a 2 (*typical* and *atypical*) × 3 (bins) two-way ANOVA which showed a main effect for both condition [$F= 137.30$, $MS= 1.269e+07$, $p < .001$] and for time bin [$F= 27066.63$, $MS= 2.502e+09$, $p < .001$] and a significant interaction between the two [$F= 37.57$, $MS= 3.473e+06$, $p < .001$].

For the space-normalized analysis, the start and end point of all trajectories was normalized to (0,0) and (1,1) respectively. The data was separated into four bins according to time (0-500 ms; 500-1000 ms; 1000-1500 ms; 1500-2000 ms) and we applied a 2 (*typical* and *atypical*) × 4 (bins) two-way ANOVA. Again, results revealed a main effect for both condition [$F= 116.20$, $MS= 24.1$, $p< .001$] and time bin [$F= 11801.39$, $MS= 2451.1$, $p< .001$], as well as a significant interaction between the two [$F= 87.71$, $MS=18.2$, $p < .001$].
Thus, we also accept the H1 of our space-normalized analysis and report significant differences in the trajectories of mouse movement between atypical and typical trials for both time-normalized and space-normalized testings.

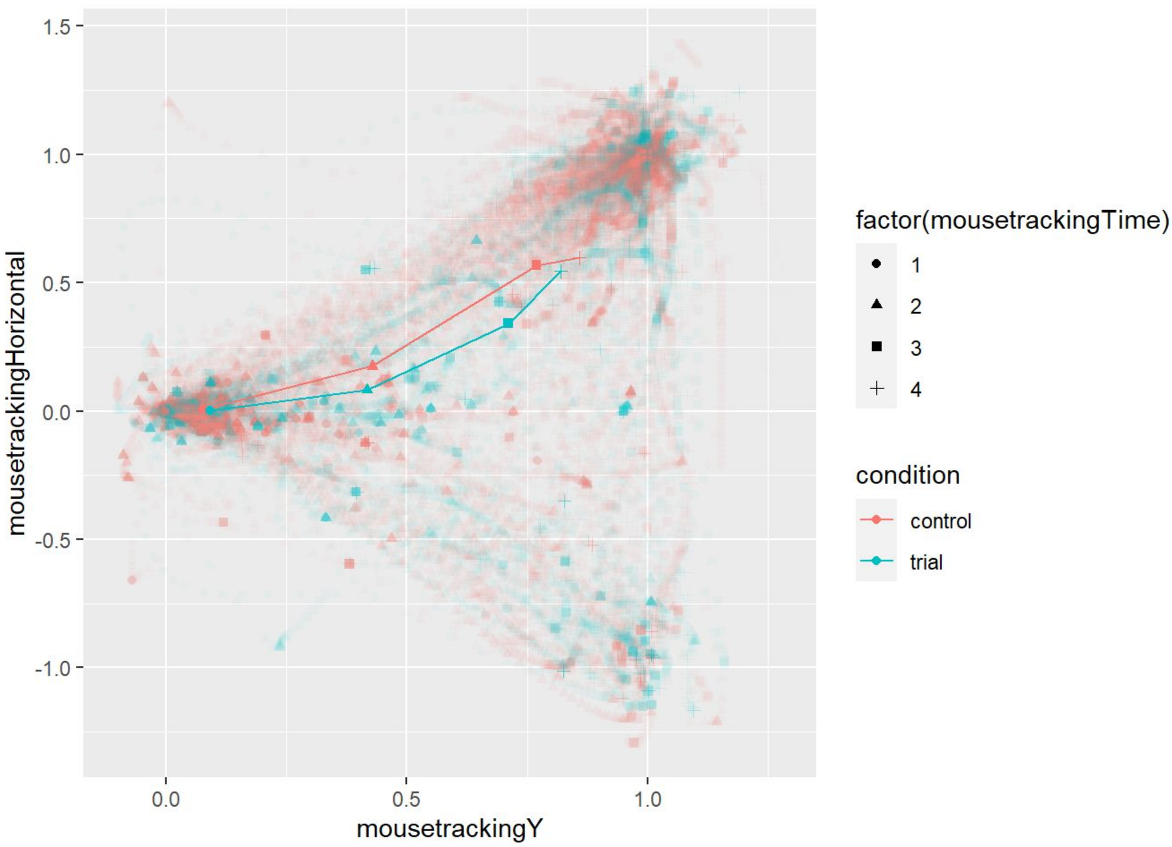

Figure 2. Mean Space-normalized trajectories of mouse movement (x-coordinates depicted on mousetrackingY and y- coordinates on mousetrackingHorizontal; data divided into four time bins indicated by four geometrical cues respectively; solid lines describe means; faint lines indicate all data samples)

Atypical trials (green line) significantly differ from typical trials (red line). Furthermore, atypical trials (green forms) show slower mouse movements than typical trials (red forms).

Discussion

Our results indicate that when comparing the mouse movement for categorization of atypical vs. typical exemplars, trajectories in the atypical condition drift significantly more to the competing category. In addition to this spatial divergence between conditions, we have also seen differences in categorization times. Looking at the trajectories, we can conclude that the categorization process evolves non-linear over time.

Considering this non-linearity, it seems reasonable to say that the manual motion of the hand reflects the underlying cognitive process, e.g. the uncertainty in categorizing atypical exemplars and its temporal dynamic.

However, a central possible weakness of this study is that the experiment was realized as a browser-based implementation: In contrast to the original study, participant's environment during the experiment could not be controlled such that there might be a higher level of distraction than in a laboratory. Therefore the categorization of atypical exemplars might have been even more difficult for the participants leading to a magnified effect. Additionally, we recruited subjects from a population in which most people's first language is german rather than english. We accounted for this by including only participants who had a sufficient level of english comprehension. However, as most participant's language skill is not comparable to that of a native speaker, this still potentially alters the timely succession of the categorization process. Especially because the questionnaire was based on self-examination in which participants' might have overestimated their own skills. Even though we found our results to support Dale's findings, we unfortunately report significantly less participants (N=29) than in the replicated study (N=41).

Another weakness of both the original study and this replicating study is a lack of consideration of possible category biases. It seems reasonable to assume that mammals can be categorized easier by humans since subjects are mammals as well (Batt, 2009). Furthermore, in our experiment as well as in the original study, participants are only included if they used the right hand to conduct the experiment. While it is usual to consider handedness in research, it leads to a certain degree of uncertainty here, since our results might just represent categorization of people who use their right hand primarily. Besides, in the original study, the mouse was placed on the right side such that participants used their right hand regardless of their actual handedness which might also have an influence on the results. One could imagine that people having to use their right hand while being left-handed can influence the trajectories of the mouse movement without reference to the categorization process.

Taking those deficits into consideration, there are multiple possibilities how future research could improve them and exploit further implications: Firstly, a german version of the here conducted experiment would be useful. As pointed out, our pool of participants might not have had the sufficient language skills and even if they had, there still might be differences in processing due to having to translate the animal name for finding the right concept, thus category. A german version of the experiment could give insight into this. Another interesting approach is a consideration of other categories and not just animals which would increase the results' generalizability.

Further, in our online version of the experiment we came across the problem of mouse tracking as a method to measure hand movement when nowadays most people use laptops, smartphones and tablets with touchpads/touchscreens to work. Use of these devices exhibits a different kind of manual movement: We suspect one will first think about the correct category and only after making a decision move the hand to click on the category. Generally, our extended use of touchscreen devices might have led to an inhibition in hand movement when making a decision. Hence, it might be interesting to investigate this further and support the experiment additionally with eye tracking data in order to not only measure the reflection of competition by means of hand movement.

Accommodating these ideas a lot of further research has already been conducted and continues to discover interesting aspects of categorization. In addition to our replication of Dale, Kehoe, & Spivey (2007) more research on the time course of categorization and mouse tracking as a method has emerged. Freeman & Ambady, for example, did not only investigate the time course of categorizing faces with the help of mouse tracking, but also analysed which conceptual features are processed first (Freeman & Ambady, 2011). This is just one example of the expanding research that contributes to our comprehension of categorization.

Reflecting on this, mouse tracking - or generally taking manual motion into account - seems to be a meaningful tool to explore categorization as a cognitive process. With research like this we can deepen our understanding of the underlying processes, its components and properties. This will hopefully lead to a more indepth theorization of conceptual knowledge and its relation to categorization which will either prove one of the many existing theories or introduce a completely new notion of knowledge representation and categorization.

Conclusion

Our experiment, despite exhibiting considerable differences, was able to replicate Dale, Kehoe, & Spivey's results. Trajectories of mouse movements in categorization processes tend to drift more to the competing category for atypical exemplars as compared to typical ones. The results implicate a non-linearity in the time course of processing, but further research is needed to investigate what that means for theoretical approaches of categorization and to account for possible weaknesses in the presented study.

References

Batt, S. (2009). Human attitudes towards animals in relation to species similarity to humans: a multivariate approach. *Bioscience horizons*, *2*(2), 180-190.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & cognition*, *35*(1), 15-28.

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in psychology*, *2*, 59.

Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin & Review*, *18*(4), 705-712.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2012). Concepts and categorization. *Handbook of Psychology, Second Edition*, *4*.