A Replication Study: Graded Motor Responses in the Time Course of Categorizing Atypical Exemplars

Julia Voss, Christina Popic, Lukas Schüler and Lars Weidtmann

University of Osnabrück

Abstract

Mouse-tracking has become a popular tool in various psychological disciplines. In the context of this paper the method is applied to investigate cognitive processes underlying the decision process in categorization tasks. One main advantage of this method being that it provides information on cognitive processes in real time, in contrast to classic reaction time paradigms. Mouse-tracking data from previous studies employing categorization tasks demonstrated stronger attraction effects towards the competing category in trials with atypical exemplars compared to trials with typical exemplars. In trials containing atypical exemplars these effects become apparent in stronger deviations in mouse movement trajectories towards the nonchosen category. This paper aims to replicate Experiment 1 from the paper "Graded motor responses in the time course of categorizing atypical exemplars" by Dale et al. (2007), as well as to test to what extent the hypotheses from the replicated paper can be generalized to stimuli belonging to other categories. For all hypotheses in our replication we found no compelling evidence or low evidence ratios and low posterior probabilities for continuous graded attraction effects of competing categories in atypical trials. The same outcomes were observed for the Capital Study.

Keywords: mouse-tracking, categorization task, cognitive processes, attraction effects, mouse movement trajectory, replication

A Replication Study: Graded Motor Responses in the Time Course of Categorizing Atypical Exemplars

The underlying mental structures of categories have long been a field of intense study in Psychology, though the most prominent theories on how categories, as collections of instances of a concept, are organized in our cognitive system have undergone a transition in the last decades. In theoretical approaches there is a move away from classical set theoretic accounts where categories possess clear boundaries and instances can only belong to a single category at a time to similarity-based fuzzy set theoretic accounts where instances can belong to one or more categories with varying degrees of how representative they are of the respective category. This inspired the development of alternative theoretic approaches such as exemplar and prototype-based theories. Amongst others the nine experiments by Rosch (1975) provided a basis of experimental evidence against classic set theoretic accounts, as these accounts could not explain why typical instances would be identified faster than less typical and why instances with a higher degree of typicality would be more affected by priming than those with a lower one. Most of the works published on the topic of mental categories concerned themselves with the outcome of the decision process to investigate the constitution of mental categories and less with the dynamics of the decision process over time. Only recently more emphasis was put on the actual time course of processing in categorization tasks. Lamberts (2000), for instance, provided an information-accumulation account to model decision processes during categorization which could accurately predict categorization response times. Besides from using reaction times, continuous measures of motor output, such as observing manual or oculomotor dynamics, can also provide valuable information about underlying cognitive processes (Freeman et al., 2011). Studies such as Cisek and Kalaska (2005) that used primates to demonstrate the intricate coupling of decision and motor processes at the level of neuron populations emphasize this point. This warrants the recording of motor action as an approach to investigate the temporal dynamics of cognitive processes. Consequently, mouse-tracking presents itself as a valuable tool in order to explore the unfolding of cognitive processes during categorization in real-time. Due to the saccadic nature of oculomotor actions mouse-tracking has the distinct advantage that it lends itself to the investigation of underlying continuous graded cognitive processes. For example, Spivey et al. (2005) discovered continuous attraction effects of mouse movement trajectories in a spoken word recognition study. Here, participants would receive spoken instructions, such as "Click on the candy!", and subsequently had to select an option from two given choices by mouse click. In trials where the choices presented were phonologically similar, just as "candle" and "candy", compared to trials where they were completely different, mouse movement trajectories showed distinct attraction effects towards the competing choice. Hence, a stronger curvature towards the competing choice disclosed the dynamic partial activation of various competing representations during the process spoken word recognition. The aim of this paper is to replicate Experiment 1 from the paper by "Graded motor responses in the time course of categorizing atypical exemplars" by Dale et al. (2007). Though instead of pursuing the frequentist approach employed in the original paper, we decided to apply Bayesian statistical inference for our analysis. Dale et al. intended to use the method of mouse-tracking to examine whether similar real-time effects of graded representations can be observed during the process of categorization of typical compared to atypical animal words. For this reason, the set of stimuli selected for the replication is comprised of highly typical and atypical animal exemplar words belonging each to one of the categories mammal, reptile, amphibian, bird, insect or fish. We also included additional trials in this replication to explore how well the original hypotheses would generalize. These additional trials were constructed utilizing typical and atypical city word stimuli. These city word stimuli had to be correctly categorized as “capital” or “non-capital”.

Method

Participants

Sample size was guided by the original paper and consequently we did not conduct a separate power analysis in order to justify sample size. We nevertheless aimed to increase the sample size from N =41, used in the original paper, to N =50 as to achieve better generalizability. In the end fifty-seven participants recruited via social media and via direct email took part in our online experiment. Only right-handed persons were considered in the final analysis. So, since three participants were left-handed and one's subject error rate was to high the final number of participants included in our analysis added up to N = 53.

Materials and Procedure

Animal and city word stimuli were presented to participants using _magpie[1] and implemented with JavaScript, CSS and HTML and hosted on Netlify[2] via GitHub[3]. Participants were given the task of selecting the appropriate category for a target word by clicking on one of two possible categories given in each trial. Target words in practice and main trials consisted of either typical or atypical animal word stimuli and of typical or atypical city word stimuli in additional trials (see Table 1). The selection of city word stimuli followed our individual, intuitive judgements about how typical they are as capitals and non-capitals of a specific country. We decided to determine the typicality of these two categories by using the degree of global popularity and the prominence of the city as the capital of its respective country as reference points. To illustrate this, Berlin would constitute a typical exemplar of the category “capital”, because it is immensely popular city and well-known to be capital of Germany. Barcelona on the other hand is not the capital of Spain, but through its considerable popularity it might present itself to participants as the apparent capital and therefore constitutes an atypical word stimulus of the category “non-capital”. We acknowledge that these intuitive judgements about typicality come along distorted by possible biases and/or lack of geographical knowledge and do not have any empirical basis. However, we regarded the city word stimuli to be sufficient to at least serve as an additional exploratory part extending our replication of the original paper.

Prior to the actual experiment participants were properly instructed to familiarize themselves with the categories given each time and to optimize speed and accuracy of their click response once the target word appeared. Moreover, we urged participants to use a computer mouse, if possible. The basic set-up of each trial was constructed according to the framework of Experiment 1 by Dale et al. (2007). In the beginning the two categories are displayed in the left and right upper corner of the screen. Following a 2000ms interval a small red button appears at the bottom of the screen, which needs to be pressed by the participant for the trial to start. When the button is pressed it disappears and the word stimulus in question appears just above the cursor. The mouse-tracking gets calibrated and the gathering of data starts until the subject decides for one category by pressing the respective button in one of the upper corners. The complete experiment contains 3 practice and 18 main trials, as well as the 14 additional trials. These additional trials serve the purpose of checking how well the hypotheses of the analysis generalize.

The main segment of the experiment is divided into two parts, practice trials and main trials. The practice trails fit the purpose of familiarizing the participant with the task. Consequently, participants are given feedback relating to the correctness of their selection. This feedback is the only aspect that differs from the structure of the main trials. To prevent the participant from adaptation to possible patterns in the sequence, in each trial type, multiple things are randomized: First the selection of the word stimulus from the dataset is randomized. Secondly, the position of the corresponding categories gets assigned randomly. And thirdly, in the case of a typical stimulus word, the selection of the competing category is randomized as well. This does not need to be done in the case of atypical word stimuli, since here the selection of categories was done by us a priori to ensure that the selection of the right category is not too unequivocal. For example, when the word stimulus is “Seal”, the corresponding categories will always be “Mammal” and “Fish”. After the main segment the participant is lead into our additional trials, divided into seven typical and seven atypical word stimuli. They constitute the same structure as the main trials but now, the only existing categories are “capital” and “non-capital".

Finally, after concluding the additional trials participants are asked to fill out a post-experiment survey including socio-demographic information and general feedback.

Our prediction, equivalent to the original paper, was that in atypical trials of the main trial block mouse movement trajectories would exhibit graded effects of competition between the categories. These effects were predicted to manifest themselves in the form a stronger curvature towards the competing category in atypical trials compared to typical trials. The same was assumed to occur in atypical trials of the additional trial block. Consequently, the only manipulated variable in this experiment is the trial type with the two levels “typical” and “atypical”. This causes trial type to be a 2-level factor with the default/reference level “atypical”. All data for the analysis was acquired via a variety of measures. The main focus of our analysis was on data obtained by recording mouse movement trajectories in x- and y-coordinates with a rate of 90 Hz compared to 40 – 80 Hz in the original paper. The second variable we recorded is the categorical variable correctness needed to describe whether the correct category was selected in each trial. This variable comes with the default / reference level “correct”. All other variables were metric measures collected in order to describe further properties of the trajectories. These variables consist of mouse movement initiation time, mouse-tracking duration, total categorization response time and distance travelled in pixels.

Table 1
Typical and Atypical Animal and City Word Stimuli for each Trial Block, With Response Options Given to the Participants (in Parentheses).

	Atypical	Typical
Practice trials	Bat (*mammal*; bird)	Lion(*mammal*)
		Rabbit(*mammal*)
Main trials	Frog (*amphibian*; reptile)	Dove (*bird*)
	Dolphin (*mammal*; fish)	Cow (*mammal*)
	Seal (*mammal*; fish)	Eagle (*bird*)
	Salamander (*amphibian*; reptile)	Tuna (*fish*)
	Sea horse (*fish*; mammal)	Fly (*insect*)
		Kangaroo (*mammal*)
		Ant (i*nsect*)
		Guinea Pig (*mammal*)
		Mosquito (*insect*)
		Zebra (*mammal*)
		Clownfish (*fish*)
		Elephant (*mammal*)
		Snake (*reptile*)
Additional trials	Barcelona (*non-capital*; **capital**)	Berlin (*capital*; **non-capital**)
	Sydney (*non-capital*; **capital**)	London (*capital*; **non-capital**)
	Beijing (*capital*; **non-capital**)	Paris (*capital*; **non-capital**)
	Ottawa (*capital*; **non-capital**)	Rome (*capital*; **non-capital**)
	Rio de Janeiro (*non-capital*; **capital**)	Moscow (*capital*; **non-capital**)
	Saigon (*non-capital*; **capital**)	Salzburg (*non-capital*; **capital**)
	Dubrovnik (*non-capital*; **capital**)	Las Vegas (*non-capital*; **capital**)

Note. Correct categories are given in Italics. As competing category labels of main, typical trials are selected in a randomized manner - either reptile, amphibian, mammal, fish, bird or insect - only correct categories are explicitly stated here. For main, atypical trials each competing category is given in boldface. For all additional trials the same typography-category mapping is applied. Italics indicate correct and boldface competing categories.

Data preparation

During the final analysis only data from correct trials was used. Moreover, data from participants who responded incorrectly to all three practice trials or had an overall accuracy of less than 50% was excluded. To correct for effects of handedness we excluded all left-handed participants. Also, all trials with categorization response times below 500ms or above 60000ms were excluded. This served to remove trials where participants had been inattentive or where participants simply clicked through the experiment quickly. For the purpose of distinguishing trajectory shapes caused by preference development from trajectory shapes emerging from other processes such as information acquisition or slips of the hand we applied a heatmap. Information acquisition might be reflected by directed movements towards a point where information was presented on the screen. Slips of the hand might be presented in erratic movements or result in movements that are unrepresentative for the context, for example, comparatively large amounts of up and down movements. In order to exclude trials with these deviations, all trials will be visually inspected in a heatmap and trials which contain more than three flips along the y-axis and the x-axis will be taken out from the analysis. Also, certain directly measured variables were combined into indices for the succeeding analysis. Mean values of the metric variables mouse movement initiation time, mouse-tracking duration and total categorization response time were computed. The complete analysis was conducted using the statistical programming language R (R Core Team, 2016) where we will rely on the 'mousetrap' package (Kieslich et al., 2019) for preprocessing, analyzing and visualizing the mouse-tracking data and the 'brms" package for analyzing Bayesian regression models. Specifically, the curvature indices *'Area under the Curve'* (AUC) and *'Maximal Absolute Deviation'* (MAD) were calculated using the 'mousetrap' package. The AUC is based on each trial's area between the actual mouse movement trajectory and a straight line originating in the position of the red button and ending at the position of the final mouse click. The

MAD is defined as the longest perpendicular distance from this straight line to a point on the actual mouse movement trajectory. Mean values of these two measures were also calculated. Other functions of the 'mousetrap' package were also used to cluster and map mouse movement trajectories (Wulff et al., 2019).

Results

The shape of individual trajectories will be assessed visually through a heatmap with mt_heatmap.

Analogous to the original paper the analysis consists of three parts, targeting the overall shape of mouse movement trajectories, their development over time and additional features of trajectories, respectively.

Time-normalized analysis

Due to variation in trial durations, we time-normalized all trajectories, so that each trajectory is represented by the same number of positions (101 by default, following Spivey et al., 2005) separated by a constant time interval. This analysis facilitated investigation of aggregated shapes of trajectories while temporal information was lost. For visual and statistical investigation all trajectories were remapped to the left and set to an equal starting point to ensure that all trajectories start and end on the same position. Then they were aggregated by trial type within participants and across subjects, resulting aggregate trajectories are displayed in Fig. 1.

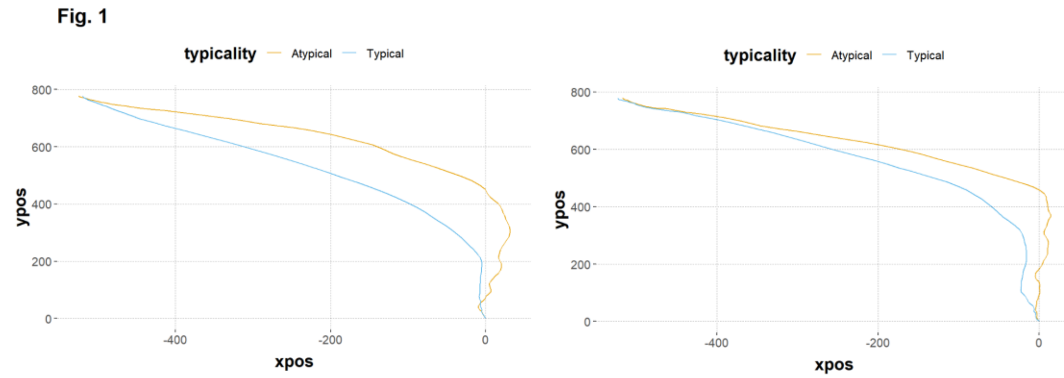

Fig. 1: Left: Display of all trials from the Original Study averaged trajectory for each trial type. Right: Display of all trials from the Capital Study averaged trajectory for each trial type. All trajectories have been flipped to the left. Starting and end point was set equal over all trajectories of each trail type

The curvature indices *'Area under the Curve'* (AUC) and *'Maximal Absolute Deviation'* (MAD), the *'Reaction Times'* (RT) and the *'Initiation Times'* will also be checked for bimodality with the 'mousetrap' function *mt_check_bimodality*. The bimodality coefficient is interpreted as bimodal for values > 0.555. None of the hypothesis's relevant measurements in the Original Study exceeded the threshold of 0.555, wherefore we can assume that there are no bimodality effects. In the Capital Study there is a bimodality effect for AUC values of the Typical trial type (BC = 0.5555211) and for the reaction times of the Typical trial type (BC = 0.5694746). For the time-normalized analysis we used the curvature indices AUC and MAD, as also described in Kieslich et al. (2020).

The time-normalized analysis was concerned with the hypothesis:

1. The area under the curve (AUC) is larger for atypical trials compared to typical trials.
2. The maximal absolute deviation (MAD) is larger for atypical trials compared to typical trials.

The credible interval of the posterior AUC values from the Original Study did not include 0. However, the evidence ratio (= 0.83) and the posterior probability (= 45%) were very small, wherefore the evidence that the hypothesis holds true is not noteworthy. The same results were observed in the Original Study for the posterior MAD values.

In the Capital Study we observed no evidence for both time-normalized hypotheses, since 0 was included in both credible intervals of the posterior AUC and MAD values.

Space-normalized analysis

In order to investigate the different trajectory types and how the development of trajectories over time depend on the experimental condition, all trajectories will be also space-normalized to 20 points each and afterwards we will extract five clusters using the *mt_cluster* function of the mousetrap R package for both trial types. The resulting cluster will be classified by prototype trajectories proposed in the meta-analysis of Wulff et al. (2019). Through assigning prototype labels to each trajectory, we get information about how the trajectories behave in space. The prototype labels are:

- *Straight*: Trajectories move directly from the start button to the chosen option
- *Curved*: Trajectories move in a curved manner from start to end point
- *Continuous Change of Mind (cCoM)*: Trajectories exhibit a curved attraction toward the nonchosen option
- *Discrete Change of Mind (dCoM)*: Trajectories move first straight to the nonchosen option and from there move horizontally to the chosen option.
- *Double Change of Mind (dCoM2):* Trajectories that first move straight to the chosen option and then horizontally switch back and forth between the nonchosen and chosen option

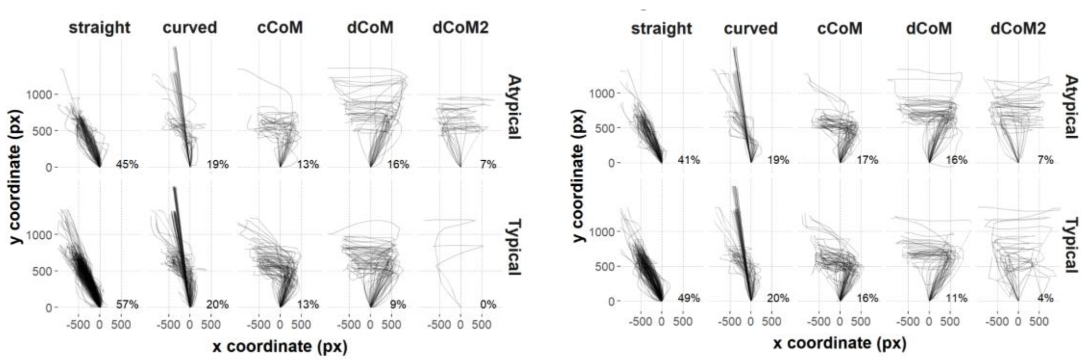

Fig. 2: The plot shows the percentages of trajectories for each trial type, which are assigned to the respective trajectory prototypes. Left: Original Study. Right: Capital Study

We account this procedure as equivalent to the space-normalized analysis procedure from the original paper (Dale et al., 2007).

The space-normalized analysis was concerned with the following hypothesis:

- The likelihood of obtaining different cluster types depends on the experimental conditions.

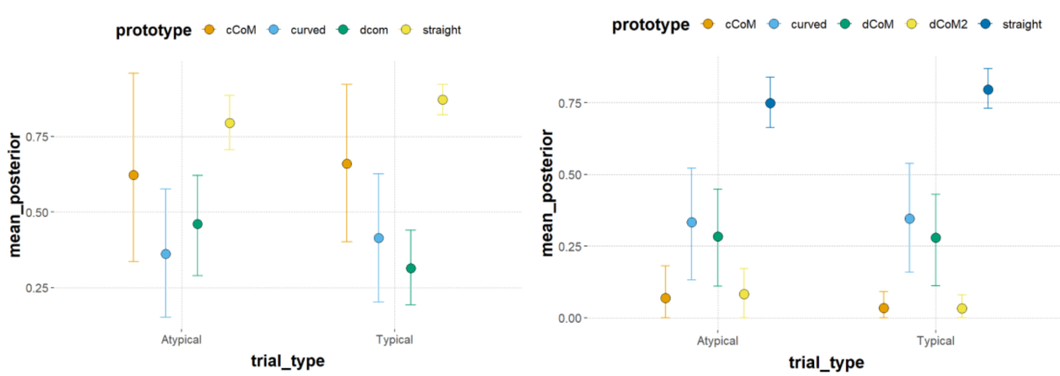

Fig. 3: The graph indicates the mean posterior probabilities of the prototypes for the Atypical and the Typical condition. Left: Original Study. Right: Capital Study

Taking from Fig.3 left there are no big differences between the prototype occurrences of trial types for the Original Study. The dCoM prototype mean posterior probability is higher for Atypical than Typical trials. Overall, no noteworthy influence of the experimental condition can be seen on the likelihood of obtaining different amounts of prototypes in the Original Study.

The right-hand side of Fig.3 shows the mean posterior probabilities of the prototypes for the respective trial types from the Capital Study. No big differences between the prototype occurrences of trial types. For the prototype straight the mean posteriors are the highest in both trial types. Overall, no striking influences of the experimental condition on obtaining different amounts of prototypes can be observed.

Additional measures

For additional measures we looked at the following hypotheses:

1. Mouse-tracking duration is longer for atypical trials compared to typical trials.
2. The total categorization response time is longer for atypical trials compared to typical trials.
3. The movement initiation latency is longer for atypical trials compared to typical trials.
4. The distance travelled in pixels is longer for atypical trials compared to typical trials

The credible interval of the posterior mouse-tracking duration values from the Original Study did not include 0. However, the evidence ratio (= 0.85) and the posterior probability (= 46%) were very small, wherefore the evidence that the hypothesis holds true is not noteworthy. The same results were observed in the Original Study for the posterior RT values and the Traveled Distance. No evidence for the hypothesis to hold true was found for the Initiation Time of the Original Study.

The credible posterior interval of the additional measures in the Capital Study did all not include 0, however the evidence ratios and the posterior probabilities were also low, which means that the evidence for the additional hypotheses of the Capital Study is barely worth mentioning.

Conclusion

Despite our appeal to all participants to make use of proper full screen mode within their internet browser, it seems many participants did not enter proper full screen mode. The problem did not occur during the pilot study. The implications of not applying proper full screen mode become apparent by visual inspection of aggregated mouse movement trajectories. Most trajectories exhibit a strong compression along the vertical dimension on the y-axis (**see Fig. 2**). Specifically, trajectories assigned to the prototypes cCoM, dCoM and dCoM2 emphasize this compression. We believe that this could have had substantial effects on mouse movement in general and consequently posed as a factor influencing the design of our experiment and the outcome of our results. Another factor we reckon to have influenced our results was whether participants used a mouse or touchpad. Due to a lack of computational power we could not control for these variables. Though both of these factors could be subjected to further analyses to explore their respective influence on mouse-tracking results. We did not observe the desired effects in the Original Study shown in the original paper (Dale et al., 2007) as well as in the additional Capital Study.

References

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*(5), 801–814. https://doi.org/10.1016/j.neuron.2005.01.027

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. https://doi.org/10.3758/bf03195938

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59. https://doi.org/10.3389/fpsyg.2011.00059

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods* (pp. 111-130). New York, NY: Routledge.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference?. *Behaviour Research Methods*, 52, 317–341. https://doi.org/10.3758/s13428-019-01228-y

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227–260. https://doi.org/10.1037/0033-295X.107.2.227

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192–233. https://doi.org/10.1037/0096-3445.104.3.192

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods* (pp. 131-145). New York, NY: Routledge.

Footnotes

[1]A minimal architecture for the creation of interactive online experiments. https://magpie-ea.github.io/magpie-site/index.html

[2]A company that offers hosting and serverless backend services for web applications and static websites. https://www.netlify.com

[3]An online platform that provides hosting for software development using the distributed version-control system Git. https://github.com