Study Information

Title

A replication study: Graded motor responses in the time course of categorizing atypical exemplars.

Authors

Lukas Schüler, Julia Voss, Christina Popic, Lars Weidtmann

Description

To extend the abundant research on static representation of category knowledge, the purpose of the original study by Dale, Kehoe & Spivey (2007) aims to investigate the time course of processing in categorization, largely unexplored up until then. The authors intend to replicate their original study in order to explore the continuous temporal dynamics of categorisation via a measure of motor output. This approach has previously proven to be informative, as studies presented evidence of graded effects over time during decision making in category judgements on motor output (McMurray et al., 2003; Nederhouser & Spivey, 2004). Further examinations of these effects could provide more information on the continuous nature of this process. In this study participants will be asked to make a category judgement - clicking on one of two categories (e.g.: *fish* or *mammal*) - after they have been presented either a typical or an atypical exemplar of one of the categories in the form of a word (e.g.: *seal* or *cow*). During this x and y coordinates of mouse movement trajectories are recorded. As preceding work (Spivey, Grosjean & Knoblich, 2005) on spoken word recognition already demonstrated attraction effects on mouse movement trajectories towards competing objects, we expect similar effects in this study. Thus we predict mouse movement trajectories to exhibit a stronger curvature towards the competing category when categorising atypical exemplars compared to when categorising typical exemplars.

Hypotheses

The authors postulate that the level of the factor “trial type” has an effect on mouse movement trajectories. To be more precise, the competitive categorization process in experimental trials that afford categorization of atypical exemplars should be reflected in the continuous mouse movements. We predict to see non-linear time course effects on mouse movement trajectories. More specifically, we expect to see a significant attraction towards the competing category on the manual motor output in experimental trials where the exemplar is an atypical animal word, compared to control conditions where the exemplar is a typical animal word. The study plans to produce evidence for these broad predictions with a variety of analyses and measures and corresponding hypotheses.

For the main trials of our experiment we hold the following research hypotheses:

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.
3. Mouse-tracking duration is longer for atypical trials compared to typical trials.
4. The likelihood of obtaining different cluster types depends on the experimental conditions.
5. The total categorisation response time is longer for atypical trials compared to typical trials.
6. The movement initiation latency is longer for atypical trials compared to typical trials.
7. The distance travelled in pixels is longer for atypical trials compared to typical trials.

Supplementing our replication hypotheses we aim to test whether these generalize. So for the additional trial block using city word stimuli we are interested in them same seven hypotheses:

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.
3. Mouse-tracking duration is longer for atypical trials compared to typical trials.
4. The likelihood of obtaining different cluster types depends on the experimental conditions.
5. The total categorisation response time is longer for atypical trials compared to typical trials.
6. The movement initiation latency is longer for atypical trials compared to typical trials.
7. The distance travelled in pixels is longer for atypical trials compared to typical trials.

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, here within the framework of an online experiment implemented using JavaScript, CSS and html and hosted on Netlify[1] via GitHub[2].

Blinding

Blinding is ensured in the form that participants are unaware of the experimental conditions they are assigned to. Furthermore, direct contact between experimenters and participants

[1] A company that offers hosting and serverless backend services for web applications and static websites. (https://www.netlify.com)
[2] An online platform that provides hosting for software development using the distributed version-control system Git. (https://github.com)

does not take place, as the experiment is conducted online. Thus any influence an experimenter could exert on participants is avoided.

Is there additional blinding in this study?

There is no additional blinding.

Study design

This study consists of a within subject design with one factor “trial type”. The factor “trial type” contains two levels, the control and experimental trial, employing typical and atypical word stimuli, respectively. The complete experiment contains 3 practice and 18 main trials, as well as 14 additional trials. All trials are presented in a random order within their respective trial block. In the practice and main trials participants are requested to judge the category of an exemplar animal word by mouse click. Within these two trial blocks the initial screen configuration in each trial displays a pairing of categories out of six possible categories: mammal, reptile, bird, fish, amphibian and insect. The two category names of each pairing are randomly assigned to either the upper left or upper right corner of the screen. For a 2000ms intervall the participants can then view the two categories, before the appearance of an additional red button in the lower middle of the screen. Following the click of the button a typical or atypical exemplar animal word is displayed at the same position. Previously, participants were instructed to then click on the category that they deem appropriate regarding the displayed animal word. For example, the initial screen would show the category name “mammal” on the upper left corner and the category name “fish” on the upper right. Then after the participant has clicked on the red button, the atypical animal exemplar word “whale” appears at the same position. As stated above, in control trials the animal exemplar words are typical category members (e.g. “cow” as a member of the category “mammal”) and category pairs presented always consist of the correct category and a competing category randomly selected from the remaining five categories (“reptile”, “amphibian”, “bird”, “insect”, “fish”). In contrast, in experimental trials animal exemplar words are atypical category members and always occur paired with the correct and the same competing category. So, an experimental trial with the animal exemplar word “seal” will always contain the category pairing of “mammal” and “fish”.
The inclusion of an additional trial block serves to see how well the hypotheses of the analysis for the main trials generalize. We selected 14 city names as city word stimuli. The stimuli were either typical or atypical capitals or cities. We adopted the same screen-setup as previously, though now participants have to categorize typical and atypical capitals and cities and category choices in each trial consist of the categories “capital” and “non-capital” while their position on the screen is randomised, being either the upper left or right corner.
The city word stimuli were selected guided by our individual, intuitive judgements about how typical they are as capitals and non-capitals of a specific country. Typicality in this sense is constituted by the degree of global popularity as a city of a specific country and by the prominence of the city to be held for the capital of their respective country. As Berlin is a popular city and similarly well-known to be the capital of Germany, this represents a typical city word stimulus of the category “capital”, whereas Ottawa is an atypical member of the category “capital”, because other major Canadian cities such as Toronto or Vancouver might be more popular and thus could be frequently mistaken as the capital of Canada. A typical

member of the “non-capital” category would be Las Vegas. It is an immensely popular city, plus it’s similarly known not to be the capital of the United States. To avoid any learning or priming effects, we avoided selecting more than one city per country.

Randomization

In the framework of this experimental within-subject design the order of trials presented within each trial block - practice, main and additional - to each participant is randomized ad hoc. Additionally, the position of categories on the left and right side of the screen is randomized and the false category is randomized for each typical trial.

Sampling Plan

Existing data

Data from a pilot study (N=5) was applied in order to guide our analysis and improve the implementation of our experimental design, if ambiguities in the description of the task or other obstacles are encountered. The preregistration takes place prior to the creation of the data used in the final analysis.

Explanation of existing data

None of the existing data from the pilot study will be used in the final analysis of the replication experiment.

Data collection procedures

Participants will be recruited through social media and via direct email. Consequently, most participants will stem from our circle of friends, classmates and family members. We acknowledge that this may present downsides to our population sample, but deem it unavoidable in the prospect of time restrictions on our experiment and the lack of financial resources.

Sample size

We aim to recruit 50 participants. In the original paper the sample size was N=41. Our slightly larger sample size takes into account that some participants might not finish the experiment and/or will be excluded according to our exclusion criteria. Given we reach our target of N = 50, this would also increase the statistical power of our analysis.

Sample size rationale

Sample size was decided on informed by the sample size used in the original paper. Seeing that our aspired sample size is larger compared to the original paper, we consider this a slight improvement due to an increased reliability of our results. Factors constraining a bigger sample size are an approaching project deadline and the lack of financial incentives for participants. Moreover, the circle of people we are able to contact via social media and email also provides an upper bound on our sample size.

Stopping rule

Collection of data will end either after having reached our target sample size or 2 days before project deadline (6th of August 2020).

Variables

Manipulated variables

The only variable that is manipulated is the trial type. The variable trial type exhibits two levels, “typical” and “atypical”, namely. Thus trial type is a 2-level factor with the default / reference level “atypical”.

Measured variables

Data for the analysis is 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. Correctness is the only categorical variable we measure. Correctness describes whether the correct category was selected in each trial or not. Therefore this is a binary variable with the default / reference level “correct”. Additionally, we take several measures that inform us about the properties of each trajectory. This includes the metric variables mouse movement initiation time, mouse-tracking duration and the total categorisation response time. Moreover, we record the metric variable distance travelled in pixels.

Indices

We will compute the mean of the metric variables mouse movement initiation time, mouse-tracking duration and the total categorisation response time. We will also create the curvature indices *‘Area under the Curve’ (AUC)* and *‘Maximal Absolute Deviation’ (MAD)*. For the AUC we will compute 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 will also be calculated.

Analysis Plan

Statistical models

Instead of pursuing the frequentist approach also employed in the original paper, we decided to apply Bayesian statistical inference for our analysis. The analysis will be written in the statistical programming language R (R Core Team, 2016), where we will rely on the ‘mousetrap’ package (Kieslich et. al., 2019) for preprocessing, analysing and visualising the

mouse-tracking data and the ‘brms’ package for analysing Bayesian regression models. We will be taken into account our beliefs through prior distributions.

Data exclusion

The final analysis will only concern itself with correct trials. We decided to correct for handedness by excluding all left-handed participants. Further, data of participants who make incorrect category judgements in all three practice trials and data of participants who only have an accuracy of 50% will not be taken into account. To prevent participants from clicking through the experiment or from being inattentive we will also exclude all trials where total categorisation response time falls below 500ms or exceeds 60000ms (equivalent to 1 minute) from the analysis. The data might contain movements that are not related to preference development but rather to other processes such as information acquisition or slips of the hand. 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. After conducting the pilot study one trial of the main trial block was excluded from the final analysis because it contained wrong category choices.

Transformations

All trajectories will be remapped to the left and set to an equal starting point to ensure that all trajectories start and end on the same position. Due to variation in trial durations, we time-normalize 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. Further, we will apply the exclusion criteria which are defined in the “Data Exclusion” paragraph. We will add curvature indices like *‘Area under the Curve’ (AUC)* and *‘Maximal Absolute Deviation’ (MAD)* for further analysis and check for bimodality with *mt_check_bimodality*. The bimodality coefficient is interpreted as bimodal for values > 0.555. Reaction times will be log-transformed. Any categorical variable such as *‘trial_type’* or *‘correct’* will be dummy-coded to 1 and 0, if needed in the analysis. The reference category will therefore always be the variable which comes first in the alphabet, i.e. the reference level for *‘trial_type’* will be *‘atypical’* and for *‘correct’* it will be *‘correct’*. The shape of individual trajectories will be assessed visually through a heatmap with mt_heatmap and compared between trial types with mt_diffmap. In order to investigate the different trajectory types and how the shapes of trajectories 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). The prototype trajectories are categorized in the following:

- 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 non chosen option
- Discrete Change of Mind (dCoM): Trajectories move first straight to the non chosen 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 non chosen and chosen option

Inference criteria

For the statistical inference, we will create 'brm' models in order to test each hypothesis. The models will be defined in the following:

H1: model_hypo1 = brm(AUC_abs_s ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with absolute, scaled AUC as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H2: model_hypo2 = brm(MAD_abs_s ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with absolute, scaled MAD as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H2: model_hypo3 = brm(mousetrackingDuration_s ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with scaled mouse tracking duration as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H4: model_hypo4 <- brm(factor_prototype_label ~ trial_type +(trial_type || subject_id)+ (1|exemplar),data = animal,family = categorical(link = "logit", refcat = NULL)

- Multinomial logistic regression model with factorized prototype_label as categorical predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H5: model_hypo5 = brm(RT_log_s ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with log transformed and scaled RT as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H6: model_hypo6 = brm(initiation_time ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with initiation time as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

H7: model_hypo7 = brm(traveledDistance ~ trial_type +(trial_type || subject_id)+(1 |exemplar)

- Mixed effect model with absolute traveled distance as continuous predictor variable, with by-subject_id varying intercepts and varying slope, without slope/intercept correlation and by-exemplar varying intercept

We will take prior beliefs into account according to the respective hypothesis. The models will provide us with useful posterior estimated coefficients from which we can report the expected values and their 95% credible intervals (CIs). In order to investigate if the hypothesis holds true or not, we will also report the posterior probability that a difference value δ, i.e. the mean of the estimated posterior for the respective coefficient, is bigger than 0. If a hypothesis states that $\delta > 0$, we can conclude that there is compelling evidence for this hypothesis if zero is not included in the 95% CI of δ and the posterior ($\delta >0$) is close to one. Model criticism will be implemented by the function 'pp_check'. In the case of comparing two models, we will be using the LOO-CV algorithm or the comparison of bayes factors.

Missing data

Additionally, in the unlikely case that some data points will not be recorded, all available data points from the participant will still be used in the analysis.

Other

References

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.

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., & Schulte-Mecklenbeck, M. (2019). Measuring the (dis-)continuous mind: What movement trajectories reveal about cognition. Manuscript in preparation