How Choice-set Configuration Affects Mouse-tracking Parameters

Chao Zhang, 26-09-2017

Introduction

Results in this report concerns the question how mouse-tracking parameters are influenced by different choice-set configurations in a food choice experiment. When we assume that *tastiness* and *healthiness* are the two major attributes participants considered when making food choices in the lab, three different types of choice-sets or trials can be distinguished (see Figure 1):

- *Dominant trials*: In a pair of food alternatives, one alternative is dominating the other one when the total utility of this alternative is much higher than the one of the other alternative. It is possible that for one of the attribute, the dominating alternative has a small disadvantage. One would hypothesize that these trials are the easist, so the associated decision processes should be fast and the mouse trajectories should be close to straight lines, with the rare exemption that the dominating alternative is not chosen.
- *Trade-off trials*: When there is no overall dominating alternative in a choice-set, trade-off describes the situation where each of the two alternatives is much more favorable than the other based on one of the two attributes, but is much inferior based on the other attribute. In these trials, a decision-maker has to sacrifice either tastiness or healthiness in making a choice. For trade-off trials, slower process is expected for resolving this conflict and mouse trajectories are likely to be more complex. Characteristics of the trajectories may also depend on whether the tastier alternative or the healthier alternative is chosen.
- *Similar trials*: When there is neither overall dominating alternatve nor trade-off, trails fall to this category. It represents the situation where two alternatives are nearly equal on both tastiness and healthiness. Decisions in this situation are expected to be very slow and mouse trajectories are also likely to be complex.

In general, analyzing the relationship between type of decision trials and mouse-tracking parameters can test our intuitions about how food choices are made. In more complex analyses (not reported here), the results may be used to evaluate different theoretical claims used in variations of dynamic decision-making models. With computational models,

decision time can be simulated and compared with the time-variables of mouse-tracking parameters (e.g., reaction time, or dragging time). In addition, momentary preferential states towards non-chosen alternatives can be simulated and compared with spatial-variables of mouse-tracking parameters (e.g., area under curve, maximal deviation, or spatial disorder).

Exploration of mouse-tracking parameters and data cleaning

At this step, six mouse-tracking parameters are examined: area under curve (AUC), maximum deviation (MD), spatial disorder (SD), dragging time (DT), maximum velocity (MV), and maximum acceleration (MA). Figure 2 plots the densities of each parameter and the relationships between pairs of them.

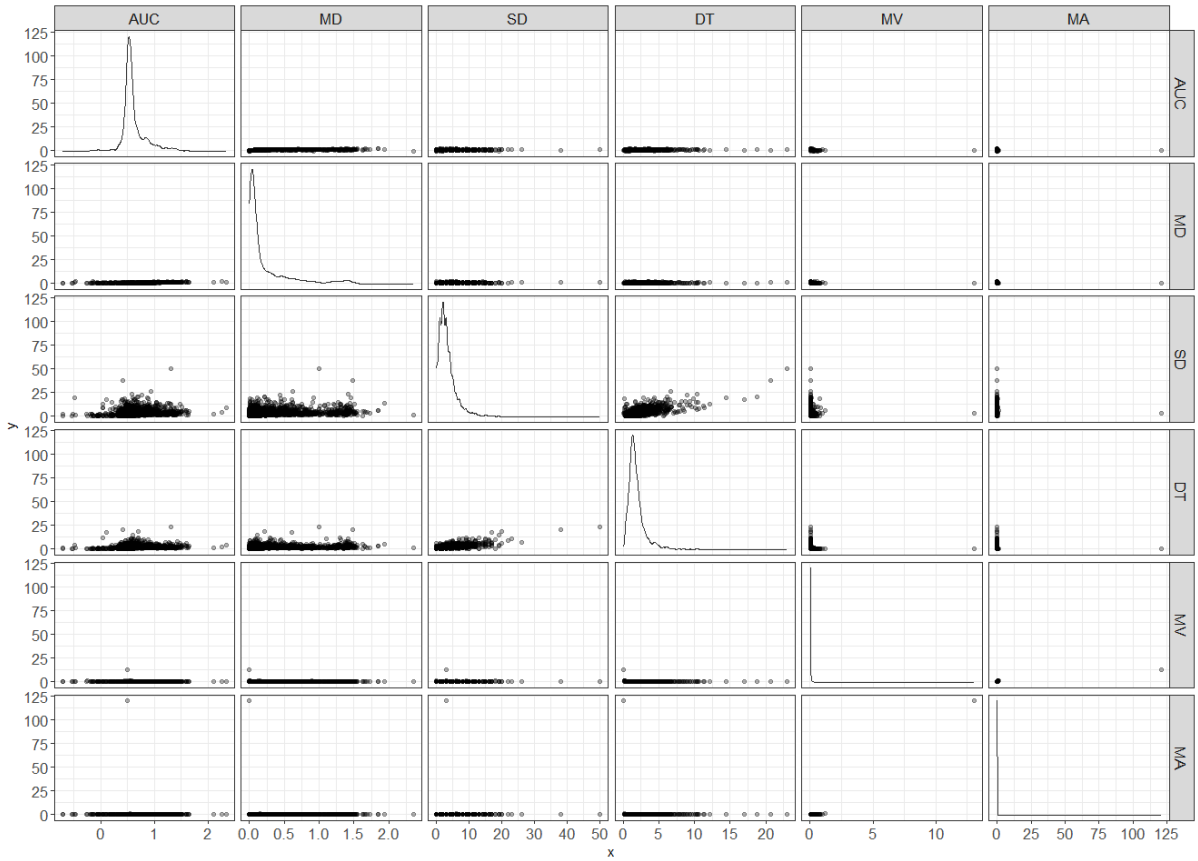

Figure 2. Density and scatterplots for raw mouse-tracking parameters

It is clear that except for AUC, the distributions of all other parameters are very skewed, and therefore log transformations were used for these variables to approach normality. During the log transformations, trials with zero MD were removed to avoid numeric problem (2.8%). There trials were probably also problematic because MD as well as AUC should be larger than zero. Further, all parameters were standardized and trials with parameter values beyond ± 3 deviations were considered as extreme values and were

removed (8.4%). Figure 3 shows the same density and scatter plots for the processed parameters.

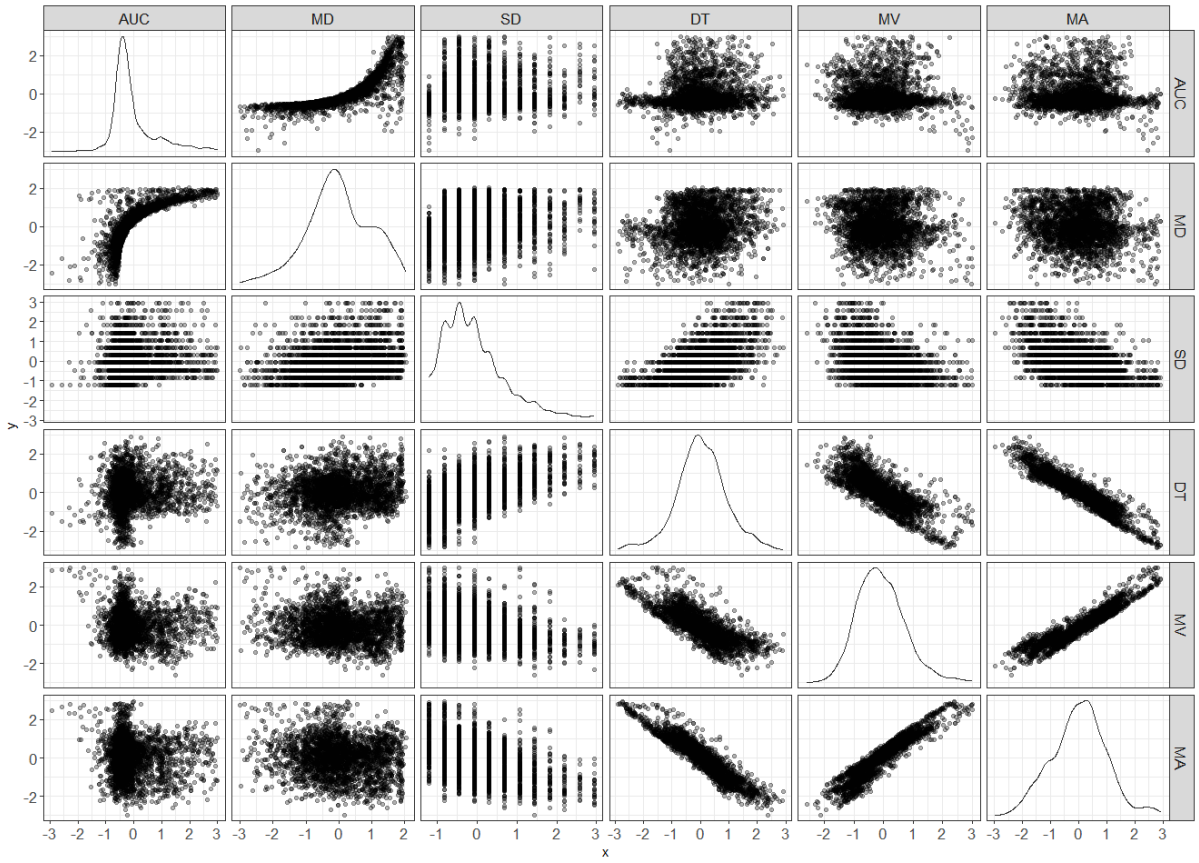

Figure 2. Density and scatterplots for processed mouse-tracking parameters

Several observations can be made based on Figure 2. First, dragging time, maximum velocity and maximum acceleration are all highly correlated. Therefore, only dragging time will be used in further analyses. Second, the distribution of AUC is very narrow, meaning that most trials had relatively small AUC within a narrow range, while a small number of "atypical" trials had large AUC. Third, the scatter plot of AUC and MD shows a curved shape (the mass), which can be explained by the constrained relationship between the two. More interestingly, in the prepheral of the mass, there were some trials with small AUC but still very large MD. These trials probably represented those that were less smooth and with abrupt deviations.

Next, in order to roughly understand the sources of variations in mouse-tracking parameters, multilevel null models were built to partition between-person and within-person variations in these variables (see Table 1). Based on intra-class correlations, AUC, maximum deviation, and dragging time all have quite large between-person variations, meaning that participants differ a lot in how they dragged the cursor on smartphone, regardless of the

decision trials faced by them and other trial-level factors. Spatial disorder, as the only ordinal variable, showed relatively small variations at between-person level.

Table 1. ICC of the parameters

	ICC
AUC	0.321
maximum deviation	0.310
spatial disorder	0.170
dragging time	0.369

Modeling choice and categorization of trials

In order to categorize each trials into the defined types, two kinds of information are required, the attribute values of the two food alternatives in a trial and the decision weights assigned to the two attributes by the decision-maker. In this study, attribute values (tastiness and healthiness) of each food item were rated by each participant after completing all the choice trials. However, individual decision weights were not measured directly and thus they need to be estimated from the data. For this purpose, a random-slope logistic regression model was built to model participants' actual choices, using difference scores of self-rated attribute values of tastiness and healthiness as predictors. One participant's data were removed from the analysis because one half of this person's trials was removed previously. This further removed 1.1% of the total number of trials.

Results showed a very good model fit, as the two measured attributes explained about 56% of the total variance (see model 1 in Table 2). Quite surprisingly, the effect of tastiness difference ($B = 1.47$, $p < 0.0001$) was more than 10 times larger than the effect of healthiness difference ($B = 0.17$, $p < 0.01$). When making the food choices, participants were mainly considering the tastiness attributes of the two food alternatives in each trial.

Since most participants had 70-90 observations, it was possible to fit one choice model for each individual participant. This procedure could be used to exclude participants who had choice data that were hardly explained by the two attributes. A poor model fit at individual level may indicate several things: (1) attributes other than tastiness and healthiness were considered during food choices; (2) the reported atttribute values from this participant were very biased from the true scores; (3) the participants made random choices for many of the trials. No matter which of three possibilities was true, excluding these participants' data was preferred for more accurate categorization of trials. Thus, participants with individual model

fit of less than 0.1 (measured as pseudo R^2) were removed (5 participants, 11.5% of total trials). Modeling results after the exclusion (model 2 in Table 2) were similar to the earlier results but with slightly improved overall model fit. The final sample for the rest of the analyses included 2463 trials from 30 participants.

Table 2. Results of modeling actual choices

	model 1	model 2
fixed effects		
intercept	-0.04 (0.05)	-0.05 (0.06)
tastiness difference	1.47 (0.17)	1.65 (0.18)
healthiness difference	0.17 (0.05)	0.18 (0.06)
random effects		
tastiness difference	0.79	0.76
healthiness difference	0.08	0.09
pseudo R^2	0.56	0.63

After fitting the model, decision weights of tastiness (WT_i) and healthiness (WH_i) for each participant were calculated as the estimated individual coefficients from the model[1]. It was interesting to observe that the estimated individual weights of healthiness did not correlate with the self-reported importance of health goal ($r = 0.13$, $p = 0.49$). Then, for each trial, utility difference between the food alternatives, and an associated trade-off score were computed using the following equations[2]:

$$Utility_L = WT_i * Tastiness_L + WH_i * Healthiness_L$$

$$Utility_R = WT_i * Tastiness_R + WH_i * Healthiness_R$$

$$\Delta Utility_{LR} = |Utility_L - Utility_R|$$

$$Tradeoff_{LR} = |WT_i * (Tastiness_L - Tastiness_R) - WH_i * (Healthiness_L - Healthiness_R)|$$

[1] The raw coefficients were used in the calculation. A different approach could to normalize the coefficients (e.g.,$wt_i = (WT_i + WH_i)$), in order to obtain the relative importance of tastiness and healthiness for each participant. However, the raw coefficients did reflect the sensitivity of each participant to the two attributes. For example, one participant could be sensitive to the changes in both tastiness and healthiness when making decisions, while another participant may be indifferent to the changes in both. As this is likely to be useful information, I decided to use the raw coefficients in the categorization of trials.

[2] As calculated in this way, utility difference scores and trade-off scores were positively correlated. Trials with high utility difference also tended to have high trade-off scores. However, only when utility difference was small (no dominating alternative), trade-off score had its meaning.

After the calculations, the following thresholds were used to categorize the trials into dominant, trade-off, and similar types:

- Dominant trials: $\Delta Utility_{LR} > 3$
- Trade-off trials: $\Delta Utility_{LR} \leq 3 \; and \; Tradeoff_{LR} > 1.5$
- Similar trials: $\Delta Utility_{LR} \leq 3 \; and \; Tradeoff_{LR} \leq 1.5$

The thresholds were arbitary, but were selected so that the number of trials in each category was balanced (so not much unlike the usually used median-split). The categorization resulted in 814 dominant trials (33.0%), 733 trade-off trials (29.8%), and 914 similar trials (37.1%). The pecentages did not add up to 1 because dominant trials in which the participants chose the non-dominate alternative were removed (only 2 trials). Figure 3 illustrates the categorization in a space of utility difference and trade-off scores.

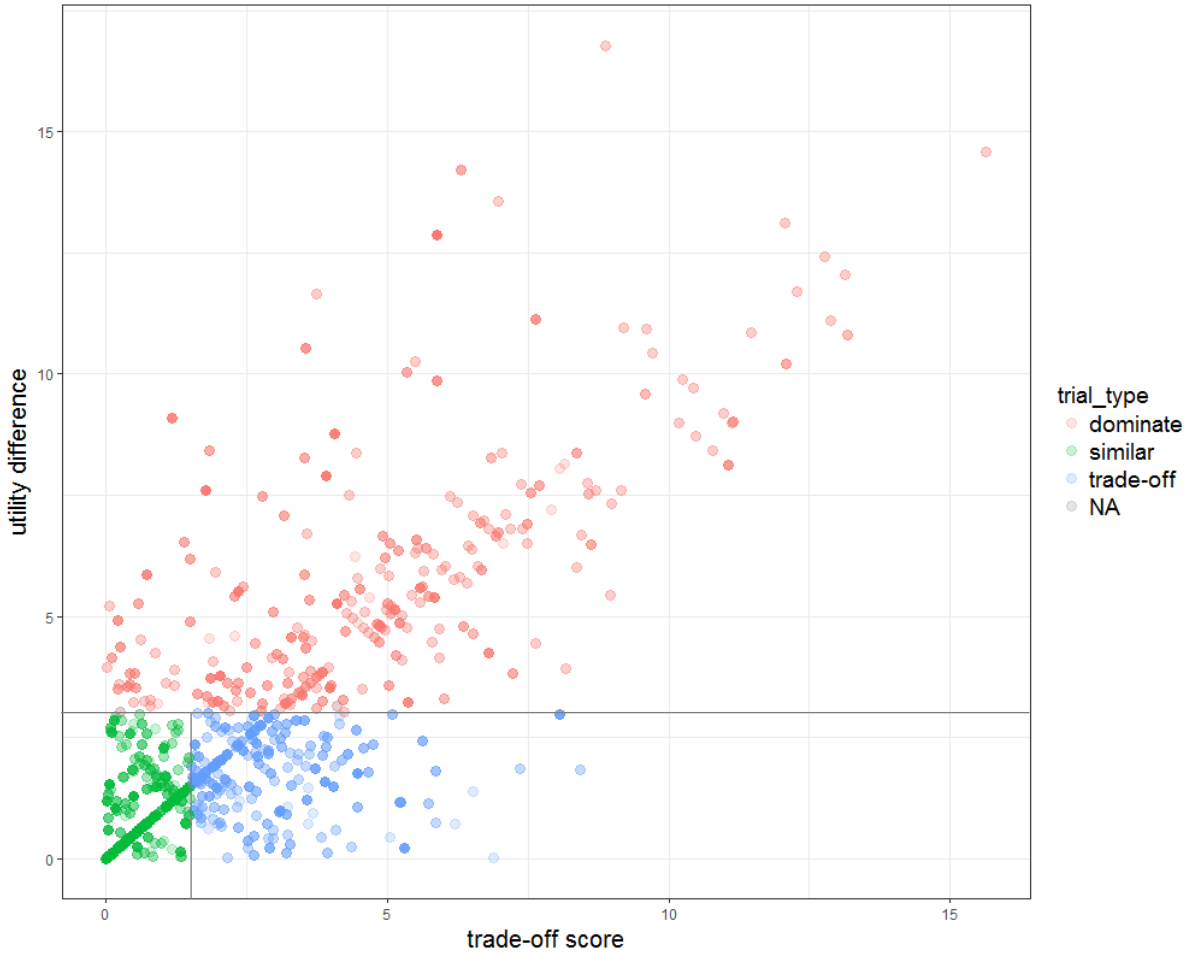

Figure 3. The categorization of decision trials

Finally, Table 3 shows 15 examples of food pairs used in the study and the percentages of them being categorized into different types of trials. The results generally met with intuitions, although there were clearly quite large individual differences in perceiving these food pairs.

Table 3. Example of food pairs as categorized

Food pair	*Domiant trial*	*Trade-off trial*	*Similar trial*
Chips - Banana	23.6%	43.6%	32.7%
Donut - Radish	34.0%	39.6%	26.4%
Fries - Pear	26.4%	39.6%	34.0%
KinderBueno - Sushi	48.1%	13.0%	38.9%
Mars - Bell pepper	40.4%	35.1%	24.6%
Chips - Fries	11.3%	32.1%	56.6%
KinderBueno - Mars	16.7%	16.7%	66.7%
Banana - Pear	10.7%	39.3%	50.0%
Bell pepper - Radish	22.2%	29.6%	48.1%
Sushi - Donuts	62.2%	15.1%	22.6%
Fries - Mars	30.0%	26.7%	43.3%
Banana - Radish	41.5%	24.5%	34.0%
Pear - Bell pepper	12.5%	26.8%	60.7%
Chips - Donuts	23.1%	26.9%	50.0%
Sushi - Bell pepper	44.6%	30.4%	25.0%

Comparing mouse-tracking parameters in different types of trials

To test whether there were robust differences in mouse-tracking parameters for the three defined trial types, one random-intercept model was built for each mouse-tracking parameters, with participant number as the grouping variable and trial type as well as choice direction (left versus right) as the predictor. For trial type, dominant trial was used as the reference level to which similar trial and trade-off trial were compared to. For choice direction, choice to the left alternative was used as the reference level. Results are summarized in Table 4. In general, the effects on all mouse-tracking parameters were very small, but given the large number of observations the effects could be reliably estimated and they were all in expected directions. In terms of sensitivity to different trial types from high to low, the parameters can be ranked in the following order: dragging time, spatial disorder,

maximum deviation, and AUC. This was a bit surprising since AUC was the most often used parameter in the mouse-tracking paradigm. Compared with dominant trials, both simlar trials and trade-off trials lead to larger AUC, maximum deviation, spatial disorder, and longer dragging time. The differences between similar trials and trade-off trials on these parameters were not significant, but some interesting trends could be noted. Trade-off trials tended to have larger AUC, but smaller spatial disorder and shorter dragging time, when compared with similar trials. Finally, choice direction seemed to have effects on AUC and the closely-related maximum deviation (actually much larger effects than the ones of trial type), but not on spatial disorder and dragging time. When choosing the food alternative on the right, participants' mouse trajectories had smaller curvatures.

Table 4. Results of modeling mouse-tracking parameters

	AUC	MD	SD	DT
fixed effects				
intercept	0.01 (0.08)	0.01 (0.10)	-0.26 (0.07)	-0.16 (0.10)
dominant trial	-	-	-	-
similar trial	0.07 (0.04)*	0.16 (0.04)***	0.27 (0.04)***	0.39 (0.04)***
trade-off trial	0.13 (0.04)***	0.18 (0.04)***	0.19 (0.04)***	0.28 (0.04)***
choice to left	-	-	-	-
choice to right	-0.18 (0.03)***	-0.13 (0.03)***	0.04 (0.03)	-0.03 (0.03)
random effects				
intercept	0.17	0.25	0.16	0.28
pseudo R^2	0.02	0.01	0.02	0.04

Note: * $p < 0.05$, *** $p < 0.001$

The effects of trial types on mouse-tracking parameters are also visually represented in Figure 4. The violin plots show the density functions of mouse-tracking parameters in each trial type, and the error bars show the means and 95% CIs of the variables. An insight from the visulization was that the peaks and centralities of the distributions of the parameters were largely the same for the three trial types, but for the more difficult trials (trade-off and similar), there were more data points in the tails on the positive side. Moreover, for dragging time, the shape of the distributions in similar trials and trade-off trials were quite different, despite that the 95% CI were largely overlapping. This pattern was interesting, but its robustness was difficult to assess.

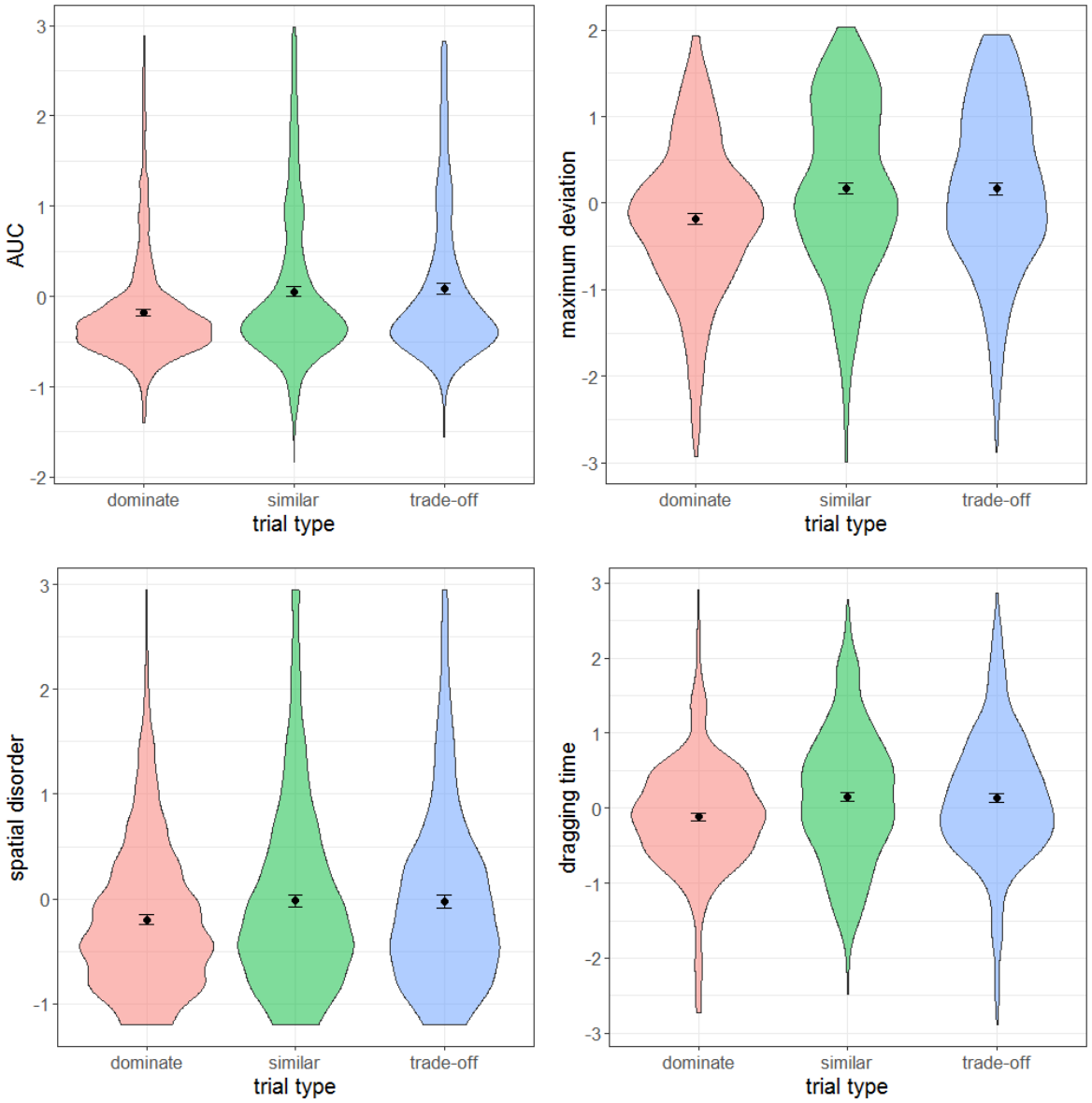

Figure 4. Violin plots of the mouse-tracking parameters against trial type

(error bars represent means and 95% CIs)