U-schape in iVR

Joern Alexander Quent

December 03, 2019

bla

# Normative data

Twenty objects were chosen from the Unity asset store and from <https://archive3d.net/>. 12 of these are typically found in a kitchen while 8 are not (see Table 1 for full list). Twenty locations were identified within the kitchen VE. In order to select where to place each object with the aim to get a range of expectancy values from unexpected, neutral to expected, a group of 6 participants were presented with screenshots of each object at each location (i.e, 400 trials in total) and asked to rate how expected that object was in that location, from -100 (unexpected) to +100 (expected). In addition, they were then asked to rate the general expectancy of each object in a kitchen (a further 20 trials). Four additional objects (kitchen: peppers and white pot; non-kitchen: dumbbells and wrench) were used to create 8 object/location practice trials, to give participants an idea about the task and calibrate their ratings. Responses were given by moving a slider across a scale. A more detailed analysis of these ratings can be found in my open, online lab-book (<https://jaquent.github.io/2018/0110_ratingAnalysis.html>). The ratings were then ranked from 1 and 400 within each participant (since different participants used different ranges internally), and the ranks were averaged over participants. Then a search algorithm was run in which one location was randomly chosen for each object on each iteration, and the iteration with the maximal spread of expectancy ranks across the 20 objects was chosen for the experiment (together with the constraint of suitable recognition foils being available; see Methods for each experiment for details).

# Experiment 1

## Population & demographics

In total, we tested 17 participants. Due to different instructions given, one participant was excluded from this analysis. Of those participants, 9 identified as female and 8 as male. Their mean age was 26.59 (SD = 3.52) years.

Due a handling error, one recall trial was not recorded for one participant. For the same participant, there is also no information available whether the participant actually saw the object (i.e. no memory trial). Nevertheless, this participants was included in this analysis because mixed linear models are robust against missing data.

## Method

The 12 kitchen objects and 8 non-kitchen objects were based on those used in Lew & Howe (2017); only objects for which no 3D model could be imported to unity3D online were replaced (Table 1). The twenty locations were the same as those in the normative study. The selection maximised the spread across the expectancy values at the same time as minimizing the difference between the target and each of the two foils used in the three-alternative forced choice (3AFC) recognition test (for full details, see XXX). Due to these tight constraints, only one set was selected, which was used for all participants. Maximising the spread was our priority because we think that previous studies did not find a U-shape relationship because they did not assay the whole range of expectedness (or schema-congruency).

## Statistical analysis

All analysis was done in R (CITE). For the main analysis, we used Bayesian mutli-level models to test our hypotheses. This analysis was done with *brms* (Bürkner, 2017, 2018), which itself is based on Stan (Carpenter et al., 2017). In all these models, individual responses for each trial are modeled as function of participants’ object/location expectancy ratings given at the end of the experiment without any from of aggregation across trials. For all logistic regression models, we scaled the data to have a mean of zero and a standard deviation (SD) of 0.5 and used Student’s t priors for the coefficients, T(*df* = 3, = 0, = 2.5), and for the intercept, T(*df* = 3, = 0, = 10), in line with Gelman et al. (2008) and [online recommendations](https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations). For those models, we used the the Bernoulli link function implemented in *bmrs*. For all models using a Gamma link function, we scaled the data to have a mean of 0 and a SD of 1 with a unit prior based derived from the standard normal distribution, N( = 0, = 1). These generic weakly informative priors are chosen to regularise unexpectedly large effects (Gelman et al., 2008). Four Markov chain Monte Carlo (MCMC) chains were run per model with 1000 warm-up and 1000 regular iterations with a total of 4000 samples per model. All of our models converged with an of 1 and the minimum number of effective samples was XXX for parameters subject to statistical test. For these analyses, evidence in favour of or against our hypotheses were quantified with Bayes factors (BF). Those BFs were calculated with the Savage-Dickey method (Wagenmakers et al. 2010). All tested comparisons were order-restricted according to our hypotheses that were pre-registered if not otherwise specified. In short, we compared the density of the truncated and renormalised prior distributions at zero with the logspline non-parametric density estimate of the truncated and renormalised posterior distributions of our parameters at zero (based on the 4000 post-warm-up samples). For unrestricted comparisons, BFs were just density ratios at zero: of *prior/posterior* (*BF10*) or *posterior/prior* (*BF01*) without truncation and renormalisation. Prior densities were estimated by using the parameters and the respective density functions (*dnorm* and *dstudent\_t*). For the analysis, we adapted code from Eric-Jan Wagenmerks’ [website](http://www.ejwagenmakers.com/papers.html). In addition to BFs, we report 95 % credible intervals around our parameters, which can be interpreted as evidence against the null hypotheses if they do not include zero. This is because Bayesian confidence interval are constructed so that the true value lies between this interval with a probability of 95 %. BFs for model comparisons between model with or without varying slopes are calculated from marginal likelihoods based on bridge sampling (see Gronau et al., 2017).

For simple mean comparisons against zero, we used Bayesian *t*-tests (Morey & Rouder, 2018) using the package ‘Bayes factor’ (version 0.9.12-4.2) with the default scale parameter of that were unrestricted.

## Results

### Overall memory performance

The average number of objects that were reported as not seen was 2.15 (SD = 1.93) in the recall task and 2.4 (SD = 1.79) in the 3AFC task. This was most commonly the knife (in half the participants), likely due to its small size.

Accuracy in the recall task was in a good range (M = 0.41, SD = 0.17) and above chance of 0.05, *BF10* = 34843. The same was true for the 3AFC task (M = 0.67, SD = 0.14), which was clearly above chance0.33, *BF10* = 223413.

### Relationship between memory performance and expectancy

#### Recall task

##### Accuracy

In the first step, we identified the appropriate model to analysis our data. In the first model, we allowed the intercepts and all slopes to vary across objects and participants, while in the second model only the intercepts varied. In a comparison we found that the second model fit the data better (*BF* = 442). Therefore, we used this model for our inferences. In this model, there was no clear linear effect, = 0.63 (95 % CI [-0.11, 1.4]), however there was large positive quadratic effect, = 1.73 (95 % CI [0.23, 3.35]), *BF10* = 8.01, as we predicted.

##### Euclidean distance

Like for recall accuracy, we identified the appropriate data to analysis our data. Since Euclidean distance is bound to zero, we used a Gamma regression model to predict memory performance. Note that because this we predicted a negative quadratic term as being consistent with our hypothesis. Comparing the varying intercept model with the model where slopes vary across objects and participants, we found that the varying intercept model fit the data better (*BF* = 1295560). Therefore, we used this model for inference. In this model, there was clear linear effect, = -0.17 (95 % CI [-0.35, -0.01]), as well as a was large negative quadratic effect, = -0.29 (95 % CI [-0.51, -0.07]), *BF10* = 8.79, as we predicted.

##### Bias towards congruent locations

If participants showed no guessing bias (see Introduction), then the expectancy of the location closest to the position they recalled (based on the normative data) should to be the same for correctly and incorrectly placed objects with both being close to zero. However, the average expectancy for incorrectly placed objects (M = 40.7, SD = 14.38) was clearly higher than for correctly placed objects (M = 2.9, SD = 10.6), *BF10* = 23914, *d* = 2.19.

#### 3AFC task

For modelling 3AFC performance, we used the same model and priors as for recall accuracy. Comparing the varying intercept model with the model where slopes vary across objects and participants, we found that the varying intercept model fit the data better (*BF* = 542). Therefore, we used this model for our inferences. In this model, there was no clear linear effect, = 0.47 (95 % CI [-0.37, 1.39]), as well as a was no clear quadratic effect, = 1 (95 % CI [-0.74, 2.74]), *BF10* = 1.1.

# Experiment 2

## Population & demographics

Five participants were excluded because they rated different stimuli. In total, 24 participants are included in the final dataset. Due to different instructions given, one participant was excluded from this analysis. Of those participants, 17 identified as female and 6 as male and 1 as non-binary. Their mean age was 24.54 (SD = 2.89) years.

## Statistical analysis

For the second experiment, we used the same approach to analyse the data. The only difference was that we used the posterior distributions of Experiment 1 where ever possible as prior distributions. That means that for logistic regression models, we estimated the family specific parameters of the Student’s t-distribution, and for Gamma regression models, we used the mean and the SD of the posterior distribution as parameters for the normal prior.

## Rationale for changes and hypotheses

Experiment 2 involved two main changes. Firstly, the twenty object-location combinations were re-selected from the normative data such that the expected expectancy for the kitchen objects included those locations with ratings close to zero. This was our priority because in Experiment 1 kitchen objects tended to be at either very expected or unexpected locations. The second change was to add a ‘remember/know’ judgment to the recall and 3AFC tasks. It is possible that the advantage of highly unexpected or highly expected locations are driven by different processes, such as recollection and familiarity respectively. Since recollection has been associated with hippocampal activity, while familiarity has been associated with cortical activity (xxx), the SLIMM model would appear to predict recollection of context for events that do not conform to a schema, but familiarity for objects that do. This is consistent with a report Lampinen2000 that atypical actions in a story recall task are associated with a higher number of remember responses, while typical actions were associated with a higher number of know responses [though see Kleider2008].

Originally, we had planned to correlate the average expectancy of an object with how often that object is remembered or judged familiar (see Supplement for original pre-registered analysis). Our predictions were that probability of remember judgement would decrease with expectancy with the reverse relationship for familiar judgements. However plotting individual trials revealed a quadratic relationship, hence we added a quadratic term to our models. To quantify evidence for remember and familiar judgements, we used unrestricted (i.e. two-sided) BFs because they do not conform with our initial hypotheses. We used the same standard priors for intercept and coefficients and model structure (varying intercepts) as for the other logistic models (see above). Familiar judgements were modeled under two assumptions: redundancy and independence. Under redundancy both trials remembered and judged familiar were coded as 1. Under independence, trials with remember judgements were excluded by coding them as NA.

## Results

### Relationship between memory performance and expectancy

#### Recall task

##### Accuracy

For this analysis we used the posterior distributions of Experiment 1 as priors for the current analysis. Therefore, we placed the following priors: intercept, T(*df* = 28.55, = -1.11, = 0.37), linear term, T(*df* = 42.73, = 0.63, = 0.37), and quadratic term, T(*df* = 44.31, = 1.73, = 0.77). The analysis revealed a clear linear term, = 1 (95 % CI [-0.74, 2.74]), and a clear quadratic term, = 0.79 (95 % CI [0.33, 1.26]), *BF10* = 28.28. The updated evidence across the first two experiments now is very high, *BF10* = 227.

##### Euclidean distance

Based on Experiment 1, we placed the following priors: intercept, N( = 0.45, = 0.18), linear term, N( = -0.17, = 0.09), and quadratic term, N( = -0.29, = 0.11). The analysis revealed a clear linear term, = -0.2 (95 % CI [-0.29, -0.1]), and a clear quadratic term, = -0.25 (95 % CI [-0.36, -0.13]), *BF10* = 6.58. The updated evidence across the first two experiments was extremely strong, *BF10* = 58.

##### Bias towards congruent locations

Like Experiment 1, the average expectancy for incorrectly-placed objects (34.66, SD = 16.92) was clearly higher than for correctly-placed objects (M = -4.61, SD = 14.89), *BF10* = 79366, *d* = 1.49.

#### 3AFC task

Based on Experiment 1, we placed the following priors: intercept, T(*df* = 14.34, = 0.85, = 0.39), linear term, T(*df* = 43.68, = 0.46, = 0.44), and quadratic term, T(*df* = 33.08, = 0.99, = 0.85). In contrast to Experiment 1, we now found a clear linear term, = 0.92 (95 % CI [0.41, 1.45]), and a clear quadratic term, = 1.49 (95 % CI [0.42, 2.59]), *BF10* = 15.61. The updated evidence across the first two experiments now is strong, *BF10* = 19.

### Relationship between expectancy and remember/familiar judgements

For remember judgements, the results showed that there is no clear linear term, = 0.11 (95 % CI [-0.5, 0.76]), *BF10* = 0.13 but in effect moderate evidence in favour of the absence of a linear term, *BF01* = 7.69. However there is a clear and very strong evidence for a qaudratic term, = 2.52 (95 % CI [0.95, 4.16]), *BF10* = 45.69. This means that both very incongruent location as well as very congruent locations are more likely to be remembered.

Both under redundancy, = -0.11 (95 % CI [-0.93, 0.72]), *BF01* = 6.31, and independence, = 0 (95 % CI [-1.04, 1.07]), *BF01* = 5.2, there was moderate evidence against a linear term. On the other hand, evidence against a quadratic term under redundancy, = 1.13 (95 % CI [-0.69, 2.99]), *BF01* = 1.42, and independence, = -0.67 (95 % CI [-2.88, 1.6]), *BF01* = 1.98, was only anecdotal.

# Experiment 3

This experiment was originally planned to increase our ability to detect that the average expectancy of an object at specific location is negatively correlated with proportion of remember judgements (see Experiment 2). For this aim, we used five sets of object/location combiniations to increase the number of data points for our regression analysis (see Supplmenent). However, we changed our analysis plan after collecting the data because averaging hid a U-shape relationship that is in fact different from that we have predicted.

## Population & demographics

One participant was excluded because they rated different stimuli. In total, 24 participants are included in the final dataset. Due to different instructions given, one participant was excluded from this analysis. Of those participants, 11 identified as female and 13 as male. Their mean age was 24.96 (SD = 3.71) years.

## Statistical analysis

We analysed the data in the same way as specified above.

## Results

### Relationship between memory performance and expectancy

#### Recall task

##### Accuracy

Based on Experiment 2, we placed the following priors: intercept, T(*df* = 31.57, = -1.47, = 0.29), linear term, T(*df* = 49.37, = 0.79, = 0.24), and quadratic term, T(*df* = 53.57, = 1.92, = 0.52). The result was that we have a clear linear term, = 0.38 (95 % CI [0.02, 0.73]), and a clear quadratic term, = 1.45 (95 % CI [0.63, 2.25]), *BF10* = 0.94. In this case, the BF slightly decreased our posterior belief that there was an quadratic relationship because it was lower than 1. However, the 95% credible intervall did not include zero and the evidence combined across three experiments now was still extreme, *BF10* = 213.

##### Euclidean distance

Based on Experiment 2, we placed the following priors: intercept, N( = 0.65, = 0.12), linear term, N( = -0.2, = 0.05), and quadratic term, N( = -0.25, = 0.06). The result was that we have a clear linear term, = -0.18 (95 % CI [-0.25, -0.09]), and a clear quadratic term, = -0.22 (95 % CI [-0.32, -0.12]) even though we had to lower our posterior believe, *BF10* = 0.22. The updated evidence across three experiments was still strong, *BF10* = 13, and the credible interval again did not include zero.

##### Bias towards congruent locations

Like Experiment 1 and 2, the average expectancy for incorrectly-placed objects (37.39, SD = 15.69) was clearly higher than for correctly-placed objects (M = 1.08, SD = 14.18), *BF10* = 662625, *d* = 1.7.

#### 3AFC task

Based on Experiment 2, we placed the following priors: intercept, T(*df* = 40.37, = 0.6, = 0.28), linear term, T(*df* = 38.49, = 0.92, = 0.26), and quadratic term, T(*df* = 55.84, = 1.49, = 0.54). The result was that we had a clear linear term, = 0.67 (95 % CI [0.31, 1.03]), and a clear quadratic term, = 1.31 (95 % CI [0.53, 2.12]), *BF10* = 2.22. The updated evidence across the first two experiments was very strong, *BF10* = 43.

### Relationship between expectancy and remember/familiar judgements

For remember judgements, we placed the following priors: intercept, T(*df* = 20.15, = -0.08, = 0.33), linear term, T(*df* = 52.16, = 0.11, = 0.32), and the quadratic term, T(*df* = 40.64, = 2.52, = 0.79) based on Experiment 2. The result was that we have no clear linear term, = -0.24 (95 % CI [-0.64, 0.18]), *BF10* = 1.13, with the cummulative evidence was still moderately in favour of null, *BF01* = 6.81, but a clear quadratic term, = 1.8 (95 % CI [0.84, 2.77]), *BF10* = 1.77. The updated evidence across Experiment 2 and 3 was very strong, *BF10* = 80.87.

For familiar judgements under redundancy, we placed the following priors: intercept, T(*df* = 32.78, = 1.83, = 0.38), linear term, T(*df* = 46.73, = -0.11, = 0.41), and quadratic term, T(*df* = 50.45, = 1.13, = 0.93) basead on Experiment 2. The result was that we have no clear linear term, = -0.36 (95 % CI [-0.88, 0.18]), *BF01* = 0.79, with the cummulative evidence being low as well, *BF01* = 4.98, as well as no clear quadratic term, = 1.11 (95 % CI [-0.13, 2.34]), *BF01* = 1.03. The updated evidence across the Experiment 2 and 3 was anecdotal, *BF01* = 1.46.

For familiar judgements under independence, we placed the following priors: intercept, T(*df* = 30.45, = 0.78, = 0.39), linear term, T(*df* = 42.68, = -0.01, = 0.51), and quadratic term, T(*df* = 32.46, = -0.67, = 1.11) based on Experiment 2. The result was that we have no clear linear term, = -0.15 (95 % CI [-0.79, 0.53]), *BF01* = 1.41, with the cummulative evidence being low as well, *BF01* = 7.33, and also no clear quadratic term, = -0.32 (95 % CI [-1.91, 1.29]), *BF01* = 1.59. The updated evidence across the Experiment 2 and 3 was moderate, *BF01* = 3.15.

# References