Chapter: boundaryVR

Joern Alexander Quent

Table of Contents

# Introduction

As we experience the world through a continuous stream of sensory input, our brains are constantly trying to predict what comes next. Prediction errors (PE) can result in “event boundaries”, which segment our memories for our experiences (add quotes). Walking into a new room is thought to trigger such a boundary, as evidenced by better temporal order memory for objects within the same room than for objects in different rooms, e.g. in a virtual environment (Horner et al., 2016). However, walking between rooms also typically results in large perceptual changes (PC). Here I report an experiment that was designed to tease apart the contributions of PE and PC to the formation of event boundaries.

I designed an “M-room” for virtual environments (add FIGURE). When traversing such a room, the viewer can only see one half of the room until they reach the middle section. This enables independent manipulation of PE and PC: PC can be induced by changing the wall colours between the two halves of the room, and PE can be induced by presenting a cue indicating the colour of the second half, which is then violated.

The first step in this study was a pilot study to verify that crossing to the second half of the room in the M-room in the absence of PC or PE does not constitute a boundary. To test this, we examined whether the superior temporal order memory for objects within the same room is similar in M-rooms and the “O-rooms” used in Horner et al. (2016). For that, participants encountered 88 objects in a series of virtual rooms.

In three experiments, I describe how I failed to replicate the boundary effect on memory (i.e. within > across) while successively removing possible confounds.

# Experiment 1

## Method

### Participant pool

In all experiments, we recruited participants from the website <https://www.prolific.co/>.

### Procedure

Participants first saw a video of another camera navigating through a series of rooms. Note that Horner et al. (2016), participants navigated through the virtual environment themselves. After watching the video, participants completed a memory task (see below for details).

### Virtual environment and stimuli

The rooms were build with SketchUp (<https://www.sketchup.com>) and then imported into unity3d (<https://unity.com/>). 88 everyday objects were downloaded (e.g. guitar, toys, household items etc.) from archive3d (<https://archive3d.net/>) and edited them either in blender (<https://www.blender.org/>) or in unity3d itself. I tried to find as many 3D versions of objects as possible that were also used in Horner et al. (2016). The object sizes were kept to be scaled realistically to the other feature of the environment.

In contrast to Horner et al., the layouts of all rooms of the same type were always identical including the positions of the tables differing only in wall colour and floor material (wood or carpet textures).

In Experiment 1, M-room and O-room alternated so that each participant saw the both types of rooms. Four videos were created where the order of the objects presented was always the same (i.e. the pencils were always presented as the first object).

Video 1 and 2 two objects were presented in the first room. The first object-object sequence was hence within-boundaries. Video 3 and 4 showed only one object in the first room. The next object was therefore presented across a spatial boundary (i.e. door). The first room in video 1 and 3 was M-shaped, while the first room in video 2 and 4 was an open plane room. The reason for this was to create four counter-balancing conditions that control for the sequences of the boundary conditions (within vs. across) and which room began the series.

Like the order of the objects, the wall colours and floor materials of the rooms were constant across the videos. The only exception to this rule was that video 3 and 4 featured 45 rooms to presented all objects. The number of wall colours (blue, brown, green, grey, orange, pink, purple, red, turquoise and yellow) and floor materials (5 different carpets and 5 different wood floors) allowed us to construct 45 unique rooms.

Both types of rooms contained three tables, however only two were used in the experiments discussed here. The first table in the room as added for future version of the experiment that I never ran. All of tables had a cardboard box placed on top of it. If the camera approached two of the tables (see labels 1 and 2 in figure above), the cardboard boxes disappeared.

While watching the video, participants in this tasks were required to judge whether an object was smaller or bigger than a reference as soon as the object appeared. In this and all subsequent versions of the Experiment the object was visible for 3 seconds. After the 3 seconds, the object disappeared and the cardboard box that was covering the object re-appeared. This controlled the time during which the object could be seen by the participant.

### Batch 1

#### Description of memory task

To asses the boundary effect, I showed participants a cue object with three images under it one of which was the target object and the other two were foils. The question the participants had to answer was “What came before this object?”.

The foils were +/- 4 positions away from the target object (see Figure). This was done to implement tighter control of the foils than it was done in Horner et al. (2016). This ensured that the two foils were always from the same room type and on the same table as the target.

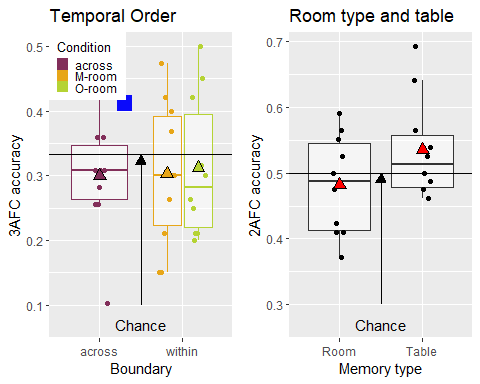
An in-lab pilot experiment has shown that participants could not identify the exact room a particular objects was in based on as still image that showed wall colour and floor texture. This memory question was therefore not used again. However, I asked participants to choose the but room type and table type in a 2AFC task.

#### Sample

In total, 10 participants (6 female and 4 male) recruited through prolific completed this version of the experiment. Their age was M = 33.3 (SD = 9.24) years.

#### Statistical analysis

#### Results



There was no conclusive evidence that memory performance in the across room condition differed as a function of whether the cue object was in an M, M = 0.329 (SD = 0.0909), or O-room, M = 0.27 (SD = 0.135), BF\_01 = 1.35, *d* = 0.477 (see Figure x), so that this condition was collapsed across rooms for further analysis.

There was moderate evidence that temporal order memory was not above chance (i.e., averaged across within and across conditions; M = 0.304 (SD = 0.0628)), BF\_01 = 6.75. This might be why there was also no evidence for a boundary effect - across, M = 0.3 (SD = 0.0921) , vs. within, M = 0.304 (SD = 0.112), - for either M rooms, BF\_01 = 3.11, *d* = -0.0179, or O-rooms, BF\_01 = 2.7, *d* = -0.0757 - across, M = 0.3 (SD = 0.0921) , vs. within, M = 0.312 (SD = 0.108).

There was evidence that participants could not remember in the room type in which an object was presented, BF\_01 = 5.07, M = 0.482 (SD = 0.0755). The evidence for whether they could remember the table on which an object was presented was anecdotal only, BF\_10 = 1.29, M = 0.536 (SD = 0.0769).

Additionally, I submitted the data to a 2 x 2 repeated measure anova with room type (O vs. M room) and boundary (across vs. within) as factors. This uncovered that there is no evidence for a main effect of room type, BF01 = 2.18, for a main effect of boundary, BF01 = 3.44 and for the interaction term, BF01 = 1.42.

#### Discussion

In contrast to my expectations, I did not find boundary effect for either room type. After consultation in a lab meeting, I considered whether the question “What came before this object?” could be interpreted in a way that participants thought that both objects that appeared before the cue object (i.e. the target as well as one of the foils) would be a valid answer.

Another notable concern was that memory performance was not above chance, which might have been another reason why I did not find the boundary effect.

### Batch 2

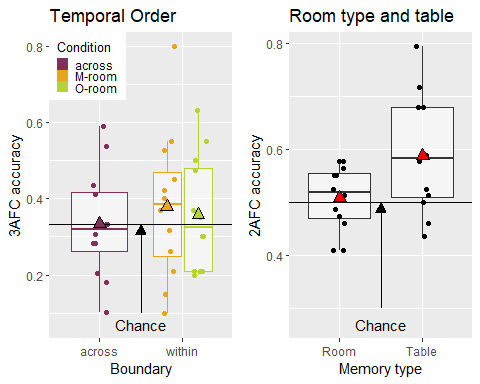
#### Description of memory task

In order to avoid confusion I improved the instructions and the specific wording of the temporal order question. Participants were now asked “In the video you just watched, which one of the three objects at the bottom of the screen appeared immediately before this object?” in order to make sure that there is no misunderstanding. Besides other small changes to the instruction of this task, no significant changes were made.

#### Sample

In total, 12 participants (5 female and 7 male) recruited through prolific completed this version of the experiment. Their age was M = 32.75 (SD = 11.02) years.

#### Results



There was again weak evidence that memory performance was not above chance, BF01 = 2.23, M = 0.351 (SD = 0.104), for the temporal memory question. Furthermore, I did not find a boundary effect for M-rooms for accuracy, across: M = 0.333 (SD = 0.143), within: M = 0.38 (SD = 0.193), BF01 = 1.59, *d* = -0.256, and I did not find an effect for O-rooms for accuracy, across: M = 0.333 (SD = 0.143), within: M = 0.359 (SD = 0.148), BF01= 2.52, *d* = -0.117.

Participants also did not show above chance performance for remembering in which room type a cue object was presented in, BF01 = 2.33, M = 0.509 (SD = 0.06). However, there was stronger evidence that participants did remember on which table type a cue object was presented, BF10 = 7.16, M = 0.589 (SD = 0.11).

Additionally, I submitted the data to a 2 x 2 repeated measure anova with room type (O vs. M room) and boundary (across vs. within) as factors. This uncovered that there is no evidence for a main effect of room type, BF01 = 1.27, for a main effect of boundary, BF01 = 3.11 and for the interaction term, BF01 = 3.06.

#### Discussion

Unfortunately, I again failed to find a boundary effect even for O-rooms. In addition, with the exception of memory for the table question, memory performance was not above chance. Once possible reason of why memory performance was so bad was that in contrast to Horner et al. (2016) I always selected foils +/-4 positions away from the target while Horner and colleagues selected random foils. Another option could be to use a more semantic task as such a pleasant/unpleasant judgement could be made, however I was concerned that most objects are fairly neutral making this task not meaningful.

### Batch 3

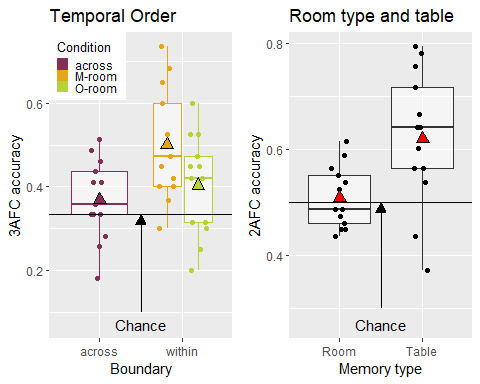
#### Sample

Due to technical problems, some prolific submissions were not usable. Unfortunately, the data containing the prolific IDs were deleted so that it not possible anymore to retrieve the demographical data only for those data that remained in the final sample of 13 participants. These came from a population of 31 (10 female and 21 male). Their mean age was M = 30.61 (SD = 7.97) years.

#### Description of memory task

Since memory performance was not above chance in batch 1 & batch 2, I decided to run another batch of participants but this time having the same foil (random) selection procedure as Horner et al (2016) with the hope that this would improve memory performance.

#### Results



This time there was strong evidence that memory performance was above chance for the temporal memory question, BF10 = 24.5, M = 0.41 (SD = 0.0778). While, I found a boundary effect for M-rooms for accuracy, across: M = 0.369 (SD = 0.0962), within: M = 0.501 (SD = 0.132), BF10 = 7.69, 0.0962, *d* = -0.774, but I did not find an effect for O-rooms for accuracy, across: M = 0.369 (SD = 0.0962), within: M = 0.404 (SD = 0.119), BF01 = 0.754, *d* = -0.43.

Interestingly, the difference between condition (across vs. within) collapsed across room type was still large, across: M = 0.369 (SD = 0.0962), within: M = 0.452 (SD = 0.0847), BF10 = 6.77, d = 0.876.

Like before, I submitted the data to a 2 x 2 repeated measure anova with room type (O vs. M room) and boundary (across vs. within) as factors. This uncovered that there is no evidence for a main effect of room type, BF01 = 3.94, for a main effect of boundary, BF01 = 0.561 and for the interaction term, BF01 = 1.47.

Participants again did not show above chance performance for remembering in which room type a cue object was presented in, BF01 = 2.15, M = 0.51 (SD = 0.0583). However, there was strong evidence that participants did remember on which table type a cue object was presented, BF10 = 19.3, M = 0.621 (SD = 0.128).

#### Discussion

Relaxing the selection of foils did have the intended effect of improving overall memory performance for temporal order. Surprisingly, I only found a boundary effect for M-room but not for O-rooms. This is surprising as with this pilot I wanted to ascertain that walking through an M-room does not have the same boundary effect that was reported in Horner et al. (2016) for walking through doors with ‘normal’ rooms. However, I found that a boundary effect for M-rooms but not for O-rooms. To investigate the absence of an boundary effect for O-rooms I designed another experiment, in which participants only saw one type of the rooms.

# Experiment 2

## Design

In this version of the task, participants were asked both temporal order question directions (Which object was before? vs. Which object was after?) in a blocked fashion. The order of which was randomised.

Table 9: Factorial design for Experiment 2

Condition

Room type

Start with

Block 1

Block 2

1

M-room

within

Before

After

2

M-room

across

Before

After

3

O-room

within

Before

After

4

O-room

across

Before

After

5

M-room

within

After

Before

6

M-room

across

After

Before

7

O-room

within

After

Before

8

O-room

across

After

Before

These 8 between-subject conditions were fully crossed: room type (M-room vs. O-room), the fist inter-item connection (within vs. across) and the direction of temporal order memory that is tested first (Which object was before? vs. Which object was after?). To keep the experiment from getting too long, room type and table type questions were scraped as sufficient evidence was found that participants can only remember on which table an object was.

Small changes were made to the lighting in the rooms making it faster to random and look more realistic. Furthermore, camera movement was now scripted instead of generated by me through navigating through the rooms myself and capturing the screen. Also made small changes were made to objects with regards to their colour and size. A few objects were also swapped because they were reported to be difficult to recognise.

## Design analysis of sequential data collection

Since I did find a difference between across and within room associations for M-rooms but not for O-rooms, I focussed on finding a boundary effect for O-rooms. To collect data efficiently, data was collected sequentially for Experiment 2. The data collection plan was as follows:

At first, I will run twelve participants for the O-room, after which I will check the evidence whether there is a difference between across and within room associations. I will increase the sample size by increments of four until we either reach a of 6 or 1/6 or the maximum sample size of 36 that is currently constrained by our remaining balance on prolific (£431.66). If I find evidence for a boundary effect for O-rooms, I will do the same for M-rooms (starting at twelve and incrementing by four) until I reach a criterion. In case, I obtain an inconclusive results for the O-rooms and I still have resources left, I will increase the sample size for O-rooms and vice versa for M-rooms.

Note that the maximal sample size of later increased to 51 (per room) as the experiment was shorter than expected.

In order to assess whether this data collection plan had enough chance to provide conclusive evidence ( > 6 | < 1/6), I ran a simplified simulation with directional Bayesian *t*-tests with minimum sample size of twelve and maximum of 36 with effect sizes d = 0, 0.44 and 0.78 that correspond to a null effect and effect sizes that we observed in Experiment 1 Batch 3.

### Null effect

### Effect size similar to boundary effect for O-rooms

### Effect size similar to boundary effect for M-rooms

### Summary

Table 13: Summary of design analysis simulation.

Effect size

Median sample size

Evidence rate

Misleading evidence rate

d = 0.00

28

0.5833

0.0301

d = 0.44

28

0.6321

0.0217

d = 0.78

12

0.9863

0.0002

In sum with exception of the third simulation, the evidence rate is not very high being around 60%. In contrast, the rate of misleading evidence was low throughout (<= 0.03). Since, the actual maximal sample size was increased after this simulation, I decided go ahead despite the relatively low evidence rate considering that I might not need a sample size of 36 for the M-room, so that I could use the remaining resources for the O-rooms.

## Planned analysis

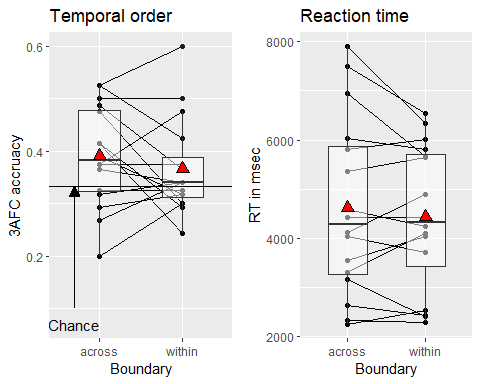
The main planned analysis was a paired t-test between within and across room associations for both room types on arcsince transformed accuracy data of the first block. I will used directional t-tests as I expect within room associations to be better than across room associations. I restricted this analysis to data from the first block because of the possibility that the second block would not yield an effect due to interference from the first block.

As additional analyses we plan to run a Bayesian ANOVA with with factors boundary (within vs across), question (Which object came before? vs. Which object came after?) and block (Block 1 and Block 2). The aim of this analysis is to find whether this is the case. In case, we find no modulation, we will also include data . Furthermore, we will run another ANOVA with boundary and room type as factors to investigate whether the boundary effect (if existing) is stronger comparing both room times. Since, this involves are between subject comparison that is not directional, we might not have sample size that is big enough. Therefore, we do not make this part of our stopping rule.

## Sample

In total, 16 participants (7 female and 9 male) recruited through prolific completed this version of the experiment. Their age was M = 27.86 (SD = 10.08) years.

## Results



Data collection terminated at N = 16, all of whom did the O-room condition, at which the BF for the null exceeded the criterion that there is no boundary effect for the first block for accuracy with O-rooms, across: M = 0.391 (SD = 0.0957), within: M = 0.366 (SD = 0.0914), BF\_01 = 7.06, *d* = -0.252. This result also held true when collapsing across both blocks, across: M = 0.398 (SD = 0.0744), within: M = 0.372 (SD = 0.0834), BF\_01 = 7.81, *d* = -0.308.

The planned ANOVA with boundary-type, question-type and block did not produce strong evidence for main effects or interactions, apart for the main effect of question type, BF\_10 = 9.23 with performance for the after question, M = 0.416 (SD = 0.0785), being higher than before question, M = 0.355 (SD = 0.0846). Overall, memory performance was low, but above chance, BF\_10 = 14.68, M = 0.385 (SD = 0.0667).

## Discussion

Unfortunately, I again failed to provide evidence for a boundary effect for O-rooms. Possible reasons for this repeated failure might be a) that memory performance might be above chance but still too low and b) rooms might look to similar to each other.

# Experiment 3

## Changes to the experiment

Surprisingly, I was not able to report evidence in favour of a boundary effect for O-rooms. However directly comparing, I found that memory performance in Horner et al (2016) was slightly higher with 0.44 across experiments and than overall memory performance was in Experiment 2, which was 0.39. The main aim for Experiment 3 was there to further improve overall memory performance.

To do this, I decided to implement three major changes a) memory encoding in Experiment 3 was intentional, b) there were two study-test cycles so that only 44 objects at a time were encoded in a block and c) rooms were made even more distinct.

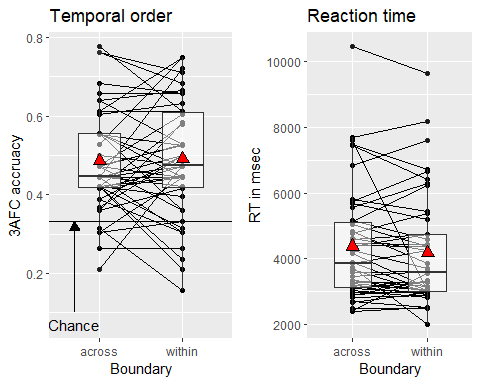
In the Experiment 1 & 2, only five different wall colours and five different floor textures were used, which means a room was only unique due to the combination fo the five colours and textures. To increase the distinctiveness in Experiment 3, each floor featured a unique texture that was selected to stand out (e.g. brightly coloured tiles, noticeable carpet patterns etc.).

## Sample

In total, 49 participants (19 female and 30 male) recruited through prolific completed this version of the experiment. Their age was M = 29.21 (SD = 11.75) years.

## Results

One participant was excluded by because they did not completed the task. Furthermore, average transformed (overall) accuracy was treated as an outlier if it was above/below two median absolute deviations from the median. This led to the exclusion of 6 % of the data. Leaving final sample of 46 participants.



Mean accuracy did increase relative to Experiment 2, up to 0.49 (0.26), which was clearly above chance BF\_10 = 1.24^{8}. However, there was no evidence for a boundary effect, in fact, some evidence against such an effect, across: M = 0.487 (SD = 0.136), within: M = 0.492 (SD = 0.147) BF\_01 = 5.44, *d* = 0.0255.

## Discussion

Even with better performance, there was still no evidence for a boundary effect of O-rooms. Due to time constraints, I have not been able to make further changes to the experimental design to replicate the basic effect.

# General discussion

In three experiments, I have been unable to provide evidence in favour of boundary effect for O-rooms and only once for M-rooms. I will therefore spent the remaining part of this chapter deliberating what could be the reasons for this.

## Potential reasons for the null effect

While I have made the rooms more distinct and the memory task itself easier successfully improving overall memory performance, two major differences between my experiment and the experiments reported in Horner et al. (2016) remain. First, in Horner et al. (2016) navigation was active, while in the series of experiment reported here participants merely watched a video. Second, the position of the tables and the floorplans in Horner et al. (2016) varied. In this experiment, rooms apart from the fact that there were M and O-rooms were the same. The room variation allowed Horner and colleagues to created a closed circle of rooms, while here the rooms all laid on a linear track.

Another difference was that I used an online sample. Even though it is conceivable that this influenced our results, it an unlikely explanation as memory performance was above chance. For instances, even in in Experiment 1 where performance was worse participant could remember on which table an object was presented.

The layout of was the rooms were in series of experiments here because it allowed to create a larger number of them and dynamically create an environment. Further options that have not been tried out that would make the room more distinct would be adding decoration to the rooms (e.g. wall paints or symbols) or even adding music or sounds for each room.