Development of Allocentric Spatial Recall from New Viewpoints in Virtual Reality

James Negen[a], Edward Heywood-Everett[a], Hannah E. Roome[a], and Marko Nardini[a]

[a]Durham University, Department of Psychology

Science Laboratories

South Road

Durham

DH1 3LE

Corresponding Author: james.negen@durham.ac.uk

edheyev@gmail.com

h.e.roome@durham.ac.uk

marko.nardini@durham.ac.uk

In press: *Developmental Science (accepted July 22nd 2016)*

Research Highlights

- Encoding spatial locations using external landmarks is a major development.
- We tested children’s abilities to recall locations in immersive virtual reality.
- One some trials, we 'teleported' them to new viewpoints.
- Children just over 4 years passed this new strict test of allocentric reasoning.

ABSTRACT

Using landmarks and other scene features to recall locations from new viewpoints is a critical skill in spatial cognition. In an immersive virtual reality task, we asked children 3.5 - 4.5 years old to remember the location of a target using various cues. On some trials they could use information from their own self-motion. On some trials they could use a view match. In the very hardest kind of trial, they were 'teleported' to a new viewpoint and could only use an allocentric spatial representation. This approach provides a strict test for allocentric coding (without either a matching viewpoint or self-motion information) while avoiding additional task demands in previous studies (does not require them to deal with a small table-top environment or to manage stronger cue conflicts). Both the younger and older groups were able to point back at the target location better than chance when they could use view matching and/or self-motion, but allocentric recall was only seen in the older group (4.0 - 4.5). In addition, we only obtained evidence for a specific kind of allocentric recall in the older group: they tracked one major axis of the space significantly above chance, $r(158) = .28$, but not the other, $r(158) = -.01$. We conclude that there is a major qualitative change in coding for spatial recall around the fourth birthday, potentially followed by further development towards fully-flexible recall from new viewpoints.

Keywords: allocentric, spatial recall, spatial development, virtual reality, early childhood

Development of Allocentric Spatial Recall from New Viewpoints in Virtual Reality

Strategies for spatial recall are categorized broadly into the realms of *egocentric* (self-centered) versus *allocentric* (world-centered). For example, if you are sitting at a desktop computer as you read this, then you could potentially encode the position of your computer mouse as down, to your right, and 40cm away from your point of view. That would be egocentric. If you wanted an allocentric version, you could instead encode the mouse's position as 25cm away from the northeast corner of your desk. Egocentric frames of reference depend on a consistent viewpoint and/or constant updating to remain accurate; if you moved your head far enough to the right, the mouse could remain 25cm from the northeast corner (allocentric remaining stable) but might end up on your left (egocentric becoming unstable). Both kinds of representation are used by adult humans and other animals (e.g. Burgess, 2006). In development, egocentric frames arise first and the move to allocentric representations is considered to be a major milestone (e.g. Piaget & Inhelder, 1956). The *strict test* of allocentric reasoning in spatial recall, as we define it here, is a spatial recall task with two strong controls on the participant's potential strategies: (1) recall must be from a new viewpoint, so that they can't just take a mental 'snapshot' and match the view, and (2) they must not have useful self-motion information between the encoding point and the recall point so that they can't just gradually update an egocentric memory as they move (e.g. Nardini et al., 2006). This test is akin to many real-world navigation tasks, such as entering a familiar market from an unfamiliar entrance and finding your way to a specific vendor.

Despite decades of research, it is still not clear when that critical transition happens (i.e. when children pass the strict test), exactly how to characterize performance just after it happens, and how this lines up against emerging performance on other spatial memory tasks. As reviewed below, previous work has either (a) allowed too many additional cues to address this question, letting participants recall from the encoding viewpoint and/or giving them useful self-motion information, or (b) imposed additional demands, using miniature table-top environments and/or creating strong cue conflicts. In this article, we take advantage of new virtual-reality technology to test children's ability to flexibly recall spatial targets after being 'teleported' to a new viewpoint in a large arena with several distinctive surrounding landmarks. This new approach avoids all of the above-mentioned issues.

Spatial Recall with a Matching View and/or Self-Motion Information

Children under two years old will pass a variety of different spatial tasks that give them more information than our defined strict test. Some of the first spatial representations are evident in neonatal orienting behaviour. Children who are only days or weeks old will turn their heads towards lights (Fantz, 1963), sounds (Clifton, Morrongiello, Kulig & Dowd, 1981), and touches (Moreau, Helfgott & Weinstein, 1978). There is also early development of *view matching*, when an organism takes a mental 'snapshot' of a place from a specific viewpoint and notices where things are by their relative appearance in that frame. Children notice differences in the arrangement of items in front of them when they are only a few months old (e.g. Wynn, 1992). A little later, we can see the use of *self-motion information*, also known as *path integration* (e.g. McNaughton et al., 2006) or *dead reckoning* (e.g. Newcombe, Huttenlocher, Drummey & Wiley, 1998), when an egocentric representation of an object's location is updated to stay in register with an organism as it moves. Infants less than a year old already show use of this skill in updating their representations of spatial layouts during simple looking-time tasks (Kaufman and Needham, 1999).

Further into development, before their second birthday, children begin successfully using self-motion information to go to remembered targets (and not just to look longer). For example, Newcombe et al. (1998) showed children a target buried in a long symmetric rectangular sandbox from one end. They were then walked to the other end of the sandbox and then allowed to try and find the target. This means that a pure view-matching process would be actively misleading; it would cause them to search near themselves at recall if the target was near them at encoding, which would be worse than even just guessing. Children as young as 16-21 months succeeded at that task despite the conflict.

Reorientation tasks show that children under 2 years can also recall locations from the same viewpoint, but without self-motion information. Disorienting a participant between encoding and recall disrupts the use of self-motion information. Children just under 2 years old will reorient themselves using environmental features in some specific circumstances (e.g. Hermer & Spelke, 1994), though seemingly-minor changes will disrupt this (e.g. Lee & Spelke, 2011). Crucially, classic reorientation tasks (e.g. Hermer & Spelke, 1994, following Cheng, 1986) involve turning the child in place, so they are released from a new facing direction but not actually a new viewpoint. In addition, they can recover both a matched facing direction and a matched viewpoint along the way to responding. This allows them in principle to use a simple 2D view matching scheme (Stürzl, Cheung, Cheng & Zeil, 2008). Interestingly, there is some reason to think that they might do something more complex anyway (Lee et al., 2012; *c.f.* Nardini et al., 2009) – but even so, a reorientation task is still not a strict test as we define it because it still does not require the child to deal with a change in viewpoint.

Spatial Recall without a Matching View or Self-Motion Information

As defined above, a strict test for allocentric recall is the ability to recall locations from new viewpoints without helpful self-motion information. A number of studies have included tests of this kind, but they have imposed additional demands on top of this already-complex task. For example, Piaget's classic three mountains task (Piaget & Inhelder, 1956) is not solved until around 9 years of age. In that task, children see a model 3D layout from one side and have to verbally describe the visible items from another viewpoint. As has been noted repeatedly (e.g. Newcombe & Huttenlocher, 1992), the task brings on additional processing beyond the central point of the task – including verbal processing, imagining a different perspective conflicting with the current perspective, and computing lines of sight. It also uses a miniature table-top layout, which can completely alter spatial behaviour (e.g. Gouteux, Vauclair &Thinus-Blanc, 2001; *see also* Learmonth et al., 2002 for another example of the effects of changing scale). The task we need to answer this question correctly is a *spatial recall* task, where we simply show them a target and ask them to point towards it after a short movement – not *navigation* (which gives additional viewpoints along the way), nor *perspective taking* (which is an additional task demand e.g. Kozhevnikov & Hegarty, 2001), nor *spatial updating* (which by definition uses self-motion cues; e.g., Wang et al., 2006).

Modern studies have followed up Piaget's work by finding evidence for competence at younger ages, but it still might be possible that additional task demands, especially the presence of strong cue conflicts, are limiting early performance. Nardini et al. (2006) provided evidence that children can pass the strict test as young as 5 years. Children were shown a toy being hidden under one of the cups in a miniature table-top sized environment and then asked to recall its location. In the crucial condition that provides a test for allocentric recall, the array was turned through 135 degrees. This disrupts the use of both self-motion information and view matching so that recall depends on encoding allocentric spatial relations on the local array of landmarks and cups. However, in this study, interpretation is made potentially difficult by a cue conflict (*see also* Nardini et al., 2009) and the environment scale. Because the spatial array is a small element of the larger environment, the task requires participants to ignore allocentric spatial relations in the larger room (e.g. looking at room features like doors and windows and using them as landmarks). A Bayesian model-based re-analysis of those data suggests that even the four-year-olds may actually have some competence at the spatial reasoning aspect of the task, but that they could not deal correctly with the cue conflict (Negen & Nardini, 2015). In the present study we take a new approach that eliminates all of these issues and so gives us more accurate insight into the development of allocentric recall.

The Present Study

The research reviewed above leads us to the yet-unanswered question of when allocentric spatial recall can be seen in a spatial recall task that meets all four of the following criteria: (1) does not give useful self-motion information, (2) requires recall from a viewpoint that is not the encoding viewpoint, (3) happens in a relatively large space, not a miniature or model layout, and (4) does not give strong cue conflicts. We reasoned that all of this was possible with immersive virtual reality technology, which allows us to generate virtual environments in which children can walk and point normally. This allows us to do something completely novel: we can show children a target within a layout and then gently 'teleport' them to a new viewpoint, which also disrupts the use of self-motion information. We can also make the virtual space look as large as we want to – we just need to constrain where participants can walk, so that they do not bump into walls in the lab. With the teleporting procedure, we do not create a conflict between the test layout and a surrounding external space, as only the test layout (in which subjects are fully immersed) is ever visible. This is also the first study, to our knowledge, to use modern immersive virtual reality with children in in this age range.

The manipulation of ‘teleporting’ participants to new viewpoints provides our major test for allocentric recall here. However, as in previous studies (e.g. Nardini et al., 2006), we also include trials in which they have simpler cues such as view matching and self-motion information available. These conditions allow us to document the transition to allocentric recall from earlier-developing representations and also serve as baselines to check that any failures are not due to misunderstanding the task.

Our aims, in summary, are two-fold: first, to carry out a strict new test of the development of allocentric coding, and second, to set this in the context of development when simpler strategies are available. To achieve these aims we use the simple task of *spatial recall* in which children see a target, experience a delay/shift, and then directly indicate its location based on different possible strategies. The task is designed to allow view matching and self-motion information on some trials and, crucially, to disrupt their use on other trials and thus require allocentric coding.

Given this new setup, which uses a naturalistic large space and reduces cue conflicts, will we see children first succeed at allocentric recall at four or five years – the same age as in a similar task (Nardini et al, 2006; Negen & Nardini, 2015) – or will they show this kind of competence earlier, nearer to the emergence of certain reorientation abilities (e.g. Hermer & Spelke, 1994)? When they do first show success on this task, will we see any evidence of an intermediate stage or any notable limits on their performance?

Method

Participants

11 children aged 3.5-4.0 years (M: 3.76, SD: 0.15), 10 children aged 4.0-4.5 years (M: 4.35, SD: 0.14), and 5 adults participated in the study. The small group of adults served primarily as a way of validating the method; their data is not used in the exact same analyses as the children and they do not contribute to any reported age effects, though we did check for expected patterns seen in other adult research. An additional 5 children aged 4.5-5.5 were also tested during piloting but are not part of the main analysis (this piloting suggested that the crucial developmental transition to allocentric reasoning is likely to be earlier). A further 3 children aged 3.0 – 3.5 did not successfully register responses (did not hold the pointer still for long enough, possibly just too young to have the required sustained attention) and are also

excluded. All the participants were recruited in and around the Durham, UK area. Children were eligible for inclusion if their parents reported no developmental concerns or disorders. Demographics were not collected. There was no paid incentive, although children were offered a small reward at the end of the study.

Apparatus

The 5m x 9m lab was fitted with Vicon Bonita cameras and motion tracking systems (Vicon, Oxford, UK). This system tracks small infrared reflective markers in real time at 240 frames per second at millimetre resolution. Once calibrated, this system allows us to track the position and rotation of objects in the room by the placement of reflective markers. A small 'wand' was made from a screwdriver handle and some pvc cylinders (see Figure 1) to act as a pointer. A hat with a stiff brim was worn by the experimenter to let the system track his location.

Display hardware. Participants viewed the virtual environment using a light-weight head-mounted display (see Figure 1), an Oculus Rift DK2 (Oculus VR, Menlo Park, CA). It was also fitted with infrared markers. This enables the participant to walk around naturally in the virtual space constrained by the cover area of the cameras. The Oculus Rift DK2 utilizes two 960 x 1080 resolution screens with refresh rates of 75 Hz. The horizontal field of view is 100°. The device itself is a lightweight headset with adjustable straps. Its size (including distance between lenses) did not appear to be a significant problem even for the three-year-olds, who were able to wear it and describe the scenery. A pilot procedure where we showed them small targets at the edges of the screen suggested that they are capable of seeing the entire field of vision. We used a default interpupillary distance of 5cm, which is 1cm less than the adult average. In two cases where the child was having trouble seeing and pointing at objects in the simulation, the problem was solved by re-measuring their interpupillary distance and adjusting the simulation accordingly.

[FIGURE 1 about here]

Environment. Figure 1 provides a screenshot of the virtual environment from overhead, another screenshot from the perspective of someone approximately 1m tall, and a photograph of a young child wearing the headset and holding the wand. The virtual environment used was a purpose-built seascape programmed in Vizard (WorldViz, Santa Barbara, CA). Everything in the virtual environment shared a 1:1 scale with actual movement in the real world. There are six important components to the environment:

(1) A 1m wide jetty forming a 135° arc on the edge of a 2.5m radius circle. This is raised 1m above the sea, providing a plausible constraint and limiting the participant’s walking area while leaving a large, open environment.

(2) A subtly textured (checker-board patterned) area of sea in the center. The texture was included to provide visual flow during walking and to help judge distance.

(3) Three landmarks around the edge of the central area, grey-brown and in the shape of rock-like structures, each visually distinct but purposefully difficult to label or distinguish semantically.

(4) The targets whose locations children had to recall were different coloured ducks which rotate on the surface of the water.

(5) The experimenter was projected as a flying red sphere with gently flapping wings that followed the experimenter's head via the motion-tracked hat. Providing this avatar for the experimenter avoided the unusual and potentially-unsettling experience of having to interact with the experimenter without seeing them. Accepting this avatar as a social entity did not appear to be a problem for the children – they variously called it a sprite, a fairy, a flying pig, a bird, a talking balloon, "Sweetie", or most commonly, the experimenter's first name.

(6) The actual 'wand' was instantiated as a virtual wand with a red line projecting out along its major axis, similar to a laser pointer. The line would turn on when the participant was being asked to point to something and turn off when they were being asked to do something else. When the red line was on and it was pointed at the water, a small white circle and cone appeared at the intersection point on the water surface. This was done to make it as easy as possible for participants to see exactly where they were pointing.

Procedure

Training and warmup trials. There were two types of training trials. The first kind of training established that participants understood and could carry out points to the duck. On each trial, a target duck appeared in a random location within 2.25m of the center of the arena and participants simply had to stand on the jetty and point down at it until a response was registered. As with all trials, the white circle at the end of the wand's red line would grow when the wand was moving rapidly and would shrink when it was being held still. A response was registered when the variation in the exact place they were pointing was very small[1] over the last 3 seconds. This was done to differentiate explorative sweeps or gestures from actual intended responses. This type of training trial was done until the participant registered a response within 50cm of the target on four consecutive trials in different locations.

The second type of training trial asked them to look at the target, wait 3 seconds, and then point to the remembered location. This was done until the participant registered four responses in a row within 1m of the target. This training established that participants understood pointing to a duck’s remembered location when it is no longer visible. Four

[1] The exact algorithm used was a weighted averaging scheme. The program took the pointing locations from the last 225 frames (3 seconds) and weighted them so that the very last frame had a weight of 0.98^0 = 1, the one just before had a weight of 0.98^1 = 0.98, the one just before that had a weight of 0.98^2 = .9604, and so on. It then took the weighted average along the x and y axis and used that as a center. It then looked at the average distance from that center (weighted the same way). The response registered when that final result was less than 0.2m. Piloting suggested that this requires holding the wand fairly still but not beyond the motor capabilities of most 3.5-4.0 year olds.

children 3.5-4.0 and one child 4.0-4.5 were excluded when they failed to progress past this point after 20 trials.

For all trials (including the data collection trials), the error distance was displayed on a separate monitor that the experimenter but not the participant could see. The participants were given very enthusiastic feedback ("Wow! That's amazing! You pointed right to the duck.") if they were correct within 0.5m, somewhat enthusiastic feedback ("Hey, good job!") if they were correct within 0.5m - 1.5m, and otherwise given negative feedback ("Hmm, he was over there" in a concerned tone). If the participant tried to point outside the arena, the experimenter refused the response ("No, silly, he can't be out there. Where was he?").

Data collection trial types. A video of an adult participant quickly completing one each of the four data collection trial types is in the supplementary material. Please note that the actual participants had higher resolution, wider field of view, and control over how they moved and where they looked. In addition, trials tended to go much slower with the young children and they tended to look around more.

[FIGURE 2 about here]

After training, there were four kinds of main data collection trials (Figure 2). All trials started with participants standing on a green cross shown on the upper side of the jetty (i.e. at the top end of the jetty in the views from above in Figures 1 and 2). From this viewpoint they saw the target (duck) appear. All types of trials were matched on the total distance that the child walked. The trial types varied the availability of view matching and self-motion information (see Figure 2). The trial type with neither of these strategies available (‘*teleport forward*’; see below) is of key interest in this study since it provides the strict test for purely-allocentric direct recall. We had children try the other trials types first so that they could be as familiar as possible with the demands of the setup when they first encountered a *teleport forward* trial.

The *same view* trials gave participants both a view match and self-motion information. The participant was shown the duck, which remained in view until they confirmed verbally that they knew where the target was. Following this, the environment faded to grey except for the jetty. The effect was as if a mist surrounded the jetty so that only the jetty was now visible. They were asked to walk halfway along the jetty to another green cross and then to return to the starting point. Then, the surrounding environment faded back into view and they were asked to point to where the duck was.

The *new view* trials gave participants self-motion information but not a view match. These trials were the same as *same view* trials, except that participants walked to the opposite end of the jetty and recalled the position of the duck from a new viewpoint.

The *teleport back* trials gave participants a view match but not self-motion information. Participants walked to the other side of the jetty (like in the *new view* trials), but they were 'teleported' back to the encoding viewpoint. This happened by the entire screen fading to black at the place participants had walked to, and then fading back up at their original starting point. This took a total of about 2.5 seconds. They were not rotated as part of the experience, but were instead just translated from one point to the other. The experience was narrated by the experimenter ("Alright, now something magical is going to happen. Wow, where are you? Look around.").

The *teleport forward* trials were the main point of doing the study. They gave participants neither self-motion information nor a view match. The only thing left was to remember something about the duck's spatial relation to the environment and use those features to find it back. This started like a *same view* trial (saw the duck, confirmed they saw it, environment went grey, walked halfway, and walked back) but then they were teleported to a new viewpoint at the opposite end of the jetty after they arrived back at the encoding viewpoint. This was also narrated for them like the *teleport back* trials.

Target zones. Actual targets were placed randomly into one of 4 zones that divided the center space of the arena. The first zone was a small circle in the centre. It was surrounded by a larger concentric circle split into 3 remaining zones. (The zones are displayed on the figures in Appendix A.) In technical terms, the first zone was any θ with $0.0m < r < 0.8m$. The next three zones were restrained to $1.0m < r < 2.25m$. They were $-11/12\pi < \theta < -5/12\pi$, $5/12\pi < \theta < 11/12\pi$, and $-3/12\pi < \theta < 3/12\pi$.

Trial and zone ordering. Following training, the main section of trials began. The trial types were split evenly into 4 blocks of 4, involving 16 trials in total. The first block proceeded always as *same view, new view, teleport back*, and then *teleport forward*. From there they were randomized within-block so that no 2 trials in a row were the same type. The zones were randomized so that each type of trial was paired with a zone exactly once but allowed for 2 trials in the same zone in a row. After that, participants completed 4 additional *teleport forward* trials (where they have the fewest cues), once in each zone in a random order. This was done to give us the highest possible statistical power in the condition where we hypothesized performance to be the closest to chance. This sums to a total of 20 data-producing trials.

Results

The raw data (all search locations) are displayed broken down by age group, condition, and target zone in Appendix A. We first simply wanted to see which age groups were above or below chance in which conditions and then took a closer look at more fine-grained distinctions in performance.

Performance vs Chance for Each Condition

We first asked which groups were above chance on which conditions, achieved using a bootstrapping method. This method was always applied to just one condition at a time, but it works for either a whole group or a single individual. We pooled all of the targets and responses together and calculated the average distance error (distance between the location pointed to and the duck's actual location). Then we shuffled the responses one million times while the targets remained static and calculated the average error for each shuffle. This gives a distribution of the average errors that would result from those targets and responses if the two had nothing to do with each other. The advantage of using this approach is that it uses the empirical distribution of searches rather than assuming what "chance" (guessing) performance would look like. For example, assuming that "chance" guessing is the average error from a uniform distribution would give a very different chance level (~2.0m) to the average error from always pointing in the middle (~1.4m). With the bootstrapping distributions calculated, we can see exactly where the actual observed average error falls and ask which conditions show error that is out on the lowest 5% of their associated distributions.

In short, all conditions were passed by all participants except for the youngest children on the *teleport forward* trials. Each of the individual adults were above chance on each condition (ps < .001), as were all five of the additional pilot children between 4.5 and 5.5 years old (ps < .05). Children aged 3.5 - 4.0 years were above chance as a group on the *same view* (p = .0006), *new view* (p = .0397), and *teleport back* (p = .0029) conditions, but not the *teleport forward* condition (p = .4695) – the strict test for allocentric recall. In other words, they were above chance when they had a view match and/or a self-motion strategy available, but not when both of those were removed. Power for this test should be >90% if the responses could even move 12.5% of the way from a randomly-chosen location towards the targets on average[2]. In contrast to the younger group, children aged 4.0-4.5 were above chance as a group in all four conditions (*p*s = .0001, .0005, .0008, .0452). These results are consistent with a transition from inability to do viewpoint-independent recall based only on visual landmarks (*teleport forward* condition) at 3.5-4.0 years to an ability to do this at 4.0-4.5 years.

After examining the data, we looked post-hoc at the two separate axes of the space in the *teleport forward* trials to gain further insight into the patterns of search and how these changed with age. As part of their strategy, it is possible that children were remembering how far the targets were towards the jetty versus towards the islands. This would predict that we can look just at the distance along the X axis of the space (which runs perpendicular to the jetty; Figure 3) and see a correlation between target and response. For the younger children, we did not see a significant correlation, $r(174) = .1561$, $p = .1463$, but we did for the older children, $r(158) = .2795$, $p = .0120$. The children could also have tried to remember how far along the jetty the targets appeared. This predicts that we should see a positive correlation along the Y axis (Figure 3). We did not see a significant correlation in either the younger group, $r(174) = -.1045$, $p = .3324$, nor the older group, $r(158) = -.0057$, $p = .96$. (All adults were each individually above chance, $p < .01$, on both axes.) This suggests that the "above chance" performance of the older group, indicative of allocentric coding, may be explained by them only remembering the target location along one major axis of the space. The lack of a correlation along either axis for the younger group again is consistent with the interpretation that they were not able to employ any allocentric strategy at all.

[2] This was examined with a short simulation study. Participants completed 88 trials in total. We ran a total of 10,000 simulations. On each simulation, we (a) drew 88 targets at random from a unit circle, (b) drew 88 responses at random from a circle with a radius of 0.5, which reflects the tighter clustering of responses than targets in the actual experiment, (c) moved each response 12.5% of the way towards its target, and (d) ran the bootstrapping analysis reported in the main text. This resulted in a *p*-value below .05 on 9,182 simulations (91.82%). As a check on the code fidelity, we confirmed that the bootstrapping analysis resulted in a p-value below .05 on only 497 (4.97%) simulations when step c was removed and the null was correct.

[FIGURE 3 about here]

The younger children also avoided the two available egocentric strategies on the *teleport forward* trials. First, for view matching: rotating all of the targets 135 degrees places them in the egocentrically-correct position in the view as the child looks directly at the center after the teleport. After this transformation, there is still no significant relation between target and response in the younger group $p = .3083$. The same holds true for every rotation from 1° to 360° in increments of 1°, no *p*s < .10. Second, for self-motion: translating all of the targets 4.6m along the Y axis places them in the place that self-motion information would cue. Again, there was no relation between target and response in the younger group, $p = .3221$, nor any translation from 10cm to 4.6m in increments of 10cm, all *p*s > .25. Instead, they pointed closer on average to the center of the arena than the actual target, $t(87) = 6.973$, $p < .001$, the view matching projection, $t(87) = 6.550$, $p < .001$, or the self-motion projection, $t(87) = 11.277$, $p < .001$.

[FIGURE 4 about here]

Factors Influencing Average Error

The top plots in Figure 4 show average error by age and condition. Each age group shows a trend where walking was more accurate than teleporting and a view match was more accurate than a new viewpoint (bars show error, so less is better). To examine the between-condition effects on average error size in adults formally, we used a 2 (view matching) x 2 (teleporting) ANOVA. Individual trials were entered so as to best lever the uneven design i.e. entering the average performance for each trial type would not allow the underlying statistical models to correctly use the fact that one of those estimates is subject to lower sampling error than the other three estimates[3]. Error was higher from a new viewpoint than the encoding viewpoint, $F(1,300) = 24.22$, $p < .0001$, but teleporting versus walking had no significant effect, $F(1,300) = .86$, $p = .3543$, nor was there an interaction between the two factors, $F(1,300) = .001$, $p = .9855$.

Next we looked at the two younger groups. For the younger participants (but not the adults), the first four *teleport forward* trials were randomized into blocks with the other three trial types, but the last four were always the last to be completed. This had no significant effect on the average error, $F(1,164) = 2.22$, $p = .1379$, nor an interaction with age group, $F(1,164) = 0.69$, $p = .4071$, so they are treated as the same for further analysis.

We scaled the raw errors to make them more comparable from child to child. We used a method similar to the bootstrap method used in the comparison with chance. For each child and each condition, we computed the possible error that they could have made with each possible combination of their targets and their responses. We then divided their actual errors by the average possible error. This means that a score of 0 would be exactly correct, a score of 1 would be exactly what you would expect by chance on average, and a score above 1 would be worse than you would expect from chance. Individual participants' scaling constants varied from 1.49m to 3.24m with a mean of 2.08m and an SD of 0.37m. The average scaled scores are shown in the bottom plots in Figure 4, in comparison with the "chance" line at 1. We analysed the between-condition and between-group effects in the younger groups as a 2 (age group) x 2 (view match) x 2 (teleporting) ANOVA on the size of scaled errors. We saw the expected main effects of view matching, $F(1,412) = 9.22$, $p = .0025$, teleporting, $F(1,412) = 9.21$, $p = .0026$, and age, $F(1,412) = 5.58$, $p = .0187$, and no significant interactions, (ps > .24).

[3] For the sake of interest and comparison: A simple t-test for the scaled error averaged over all trials within each participant shows an age effect, $t(19)=2.38$, $p=.0140$. We also averaged scaled error for each condition within each subject and ran a mixed-model ANOVA. This showed an effect of view matching, $F(1,19)=6.945$, $p=.016$, and teleporting, $F(1,19)=89.637$, $p<.001$, with non-significant interactions, ps>.25.

Going back to the question of age-related performance differences specifically on the critical *teleport forward* trials, we also ran a planned comparison between age groups on just that trial type and found significantly-lower scaled error for the older group, $t(19) = 2.11$, $p = .0242$. This suggests that the difference between the significant and non-significant findings in terms of comparison against chance reflect an actual difference in average performance on *teleport forward* trials, which reinforces the interpretation that some meaningful shift tends to happen around 4.0 years in terms of allocentric reasoning in spatial recall.

Discussion

Our central aim in this project was to re-assess when children pass the strict test of allocentric reasoning, namely spatial recall from a new viewpoint without the aid of helpful self-motion information. This is a skill that is used routinely by adult navigators e.g. when entering familiar places from new entrances. We studied this by showing children a duck in a virtual duck pond, moving them in different ways, and then having them try to point to where the duck was. Sometimes the recall viewpoint was the same as the encoding viewpoint, allowing them to use a view matching strategy, and sometimes they ended up at a new viewpoint at recall. Sometimes they walked to the recall point, allowing them to use self-motion information, and sometimes they were 'teleported' there. In the specific *teleport forward* condition, they were teleported forward to a new viewpoint. This final condition meets our definition of a strict test of allocentric reasoning but does not introduce the additional demands seen in previous research (e.g. strong cue conflicts in Nardini et al., 2006, 2009).

Our 2 major findings are that (1) the group of children just under 4 years performed above chance when they had either a matching view or self-motion information, or both, but not without those extra cues, and (2) the group of children just over 4 years performed above chance (and significantly better than the younger children) even without those extra cues, though we only saw significant evidence for remembering where the target was along 1 major axis of the space. There was an additional minor finding: accuracy improved with the 2 extra cues overall for the children, but adults were only affected by a change in view – teleporting them to a different location had the same impact on average error as just having them walk there.

This leads us to 2 theoretical conclusions. First, children can pass the strict test at about 4.0-4.5 years old in a group-level analysis. Some level of inter-individual variation almost certainly must exist, but that is the point when it can be seen clearly in a larger group. This is long after succeeding in some reorientation paradigms without a change in view (e.g. Hermer & Spelke, 1994; *see also* Stürzl et al., 2008) and before passing paradigms with stronger cue conflicts (e.g. Nardini et al., 2006; 2009).

Executive function could be a key reason why we found evidence of children passing the strict test before other studies (e.g. Nardini et al., 2006; 2009). Selecting the correct representation instead of a competing incorrect one is likely to require skills related to inhibition, which, together with other aspects of executive function, develops markedly over this age range (e.g., Garon, Bryson & Smith, 2008). For example, Nardini et al. (2006) required children to ignore a global allocentric frame in the larger testing room and attend to a local allocentric frame on a smaller array. It is possible that the 4-year-olds in that study could compute the correct local allocentric frame but did not have the inhibition skills needed to let that computation control behaviour (*see also* Negen & Nardini, 2015). Even in the present study, children were still asked to ignore possible egocentric responses, and it is possible that individual differences in executive function could predict performance on our task. However, further research is needed to clarify this relation.

We can further compare our results to those from a paradigm that also has a change in viewpoint but uses a disorientation procedure in place of a teleport (Nardini et al., 2009).

Both seem *a priori* like reasonable ways of preventing the use of self-motion. While we found evidence of allocentric reasoning in 4-year-olds, they found that disorienting a 4-year-old actually makes them revert back to just view matching. It is possible that a full disorientation procedure disrupts the recall of an allocentric memory and that the change in viewpoint is enough already to cue children at this age that they need to ignore self-motion.

Crucially, the overall performance of the younger group cannot be explained readily by simple confusion in the *teleport forward* trials; children aged 3.5 – 4.0 years were capable of performing above chance on trials with both potentially-confusing elements of the *teleport forward* trials. In the *new view* trials, children of this age were able to recognize that an attempt at a view match would be misleading and respond above chance. In the *teleport back* trials, they were able to recognize that the self-motion information would be misleading and respond above chance. In the *teleport forward* trials, their responses were not significantly related to either the self-motion projection or the view-matching projection. Instead, the responses are nearer the center of the arena (versus the targets or either ego projection) on average. They look to have chosen a strategy that ignores the two egocentric cues and just defaults to the single fixed response point that is going to be nearest to the most targets. This finding stands in contrast to the many studies that have observed actively-incorrect egocentric behaviour on various spatial tasks throughout infancy and childhood when participants are simply confused (e.g. Acredolo, 1978; Nardini et al., 2006, 2009; Piaget & Inhelder, 1956). This suggests they understood a great deal about the situation – except how to use the successful allocentric strategy that the older children employed.

Of course, there are still open questions about the exact nature of the deficit that the 3.5-4.0 year olds are showing. For example, it might be possible to approach the *teleport forward* trials by estimating the rotation of the teleport around the center of the arena and trying to rotate their spatial memory by the same amount about the same point. In that scenario, the estimate of the teleport rotation, and/or the attempt to rotate their memory, and/or the identification of the appropriate center of rotation could be the issue. For another example, it might be possible to try and remember the target's location along the space's principal axis. In that case, performance could break down if the space's estimated principal axis does not match from the two viewpoints, and/or their estimate of distance along it is not consistent from the two viewpoints. There are many such possibilities that will require further research to disentangle.

As for the second major conclusion, when allocentric spatial recall first emerges on the present task it may be limited and may not track all relevant aspects of a space. Our testing arena had 2 major axes, one perpendicular to the jetty (X axis) and one along the length of the jetty (Y axis). Children just over 4 years gave responses that were significantly correlated with the targets along the X axis but not the Y axis; they remembered where the targets were in terms of how far *from* the jetty they were but did not show any evidence of remembering the targets in terms of how far *along* the jetty they were. (Adults remembered both axes.) This suggests that there are intermediate stages in the use of allocentric information in spatial recall tasks. The choice of the X axis over the Y axis is likely just due to the environment specifics; the two mountains whose locations only differ by Y coordinates were not symmetric, but they were not as different as the walkway and mountain whose locations only differed by X coordinates. This likely made the X axis more salient.

The adults were well above chance in all conditions and were affected by a new viewpoint, but not by the disruption of self-motion information, which is generally not the pattern of results seen in table-top studies of adult spatial cognition (e.g. Burgess, Spiers & Paleologou, 2004; Simons & Wang, 1998; Wang & Simons, 1999). It's possible that this is a simple ceiling effect or that the path is too long (~7m) to provide self-motion information that is measurably useful for adults alongside the much more accurate visual cues that are also available. In any case, our main interest in collecting the adult data was to be sure that it was possible in principle to be extremely accurate on this task with this equipment, which was true.

In our future work we are going to extend these studies later into development with tasks that are designed to distinguish simpler and more complex forms of allocentric spatial recall. We have made a first foray towards this goal by looking separately at the X and Y axes of the responses in this task but there is a great deal more room to explore. Another important finding from these data is the simple fact that sensible spatial recall data can be gathered with children as young as 3.5 years old with modern immersive virtual reality, even in simulations where we gently bend the rules of space and time. This opens up many possibilities for future exploration of children's spatial skills.

In summary, if you hold children the strict standard of spatial recall from a new viewpoint without helpful self-motion information, they struggle to move beyond egocentric representations until surprisingly late, just after their fourth birthday (at least in a group-level analysis) – and even then, they may just remember a single crucial aspect of the environment such as a single major axis. In this specific sense, it may take long past the fourth birthday to actually begin remembering a space in a way that is fully allocentric and qualitatively adult-like.

References

Acredolo, L. P. (1978). Development of spatial orientation in infancy. *Developmental Psychology*, *14*(3), 224.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in cognitive sciences*, *10*(12), 551-557.

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition*, *94*(2), 149-166.

Clifton, R. K., Morrongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Newborns' orientation toward sound: Possible implications for cortical development. *Child Development*, 833-838.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*(2), 149-178.

Cheng, K., Huttenlocher, J., & Newcombe, N. S. (2013). 25 years of research on the use of geometry in spatial reorientation: a current theoretical perspective. *Psychonomic bulletin & review*, *20*(6), 1033-1054.

Fantz, R. L. (1963). Pattern vision in newborn infants. *Science*, *140*(3564), 296-297.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin*, *134*(1), 31.

Gouteux, S., Vauclair, J., & Thinus-Blanc, C. (2001). Reorientation in a small-scale environment by 3-, 4-, and 5-year-old children. *Cognitive Development*, *16*(3), 853-869.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*(6484), 57-59.

Kaufman, J., & Needham, A. (1999). Objective spatial coding by 6.5-month-old infants in a visual dishabituation task. *Developmental Science*, *2*(4), 432-441.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), 745-756.

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children's use of landmarks: Implications for modularity theory. *Psychological Science*, *13*(4), 337-341.

Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddlers' use of metric information and landmarks to reorient. *Journal of experimental child psychology*, *80*(3), 225-244.

Lee, S. A., Winkler-Rhoades, N., & Spelke, E. S. (2012). Spontaneous reorientation is guided by perceived surface distance, not by image matching or comparison. *PLOS ONE*, DOI: 10.1371/journal.pone.0051373.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663-678.

Moreau, T., Helfgott, E., Weinstein, P., & Milner, P. (1978). Lateral differences in habituation of ipsilateral head-turning to repeated tactile stimulation in the human newborn. *Perceptual and motor skills*, *46*(2), 427-436.

Nardini, M., Burgess, N., Breckenridge, K., & Atkinson, J. (2006). Differential developmental trajectories for egocentric, environmental and intrinsic frames of reference in spatial memory. *Cognition*, *101*(1), 153-172.

Nardini, M., Thomas, R. L., Knowland, V. C., Braddick, O. J., & Atkinson, J. (2009). A viewpoint-independent process for spatial reorientation. *Cognition*,*112*(2), 241-248.

Negen, J., & Nardini, M. (2015). Four-Year-Olds Use a Mixture of Spatial Reference Frames. *PloS one*, *10*(7), e0131984.

Newcombe, N., & Huttenlocher, J. (1992). Children's early ability to solve perspective-taking problems. *Developmental Psychology*, *28*(4), 635.

Newcombe, N., Huttenlocher, J., Drummey, A. B., & Wiley, J. G. (1998). The development of spatial location coding: Place learning and dead reckoning in the second and third years. *Cognitive Development*, *13*(2), 185-200.

Piaget, J., & Inhelder, B. (1956). The child’s conception of space. *New York: Humanities Pr*.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*(4), 315-320.

Stürzl, W., Cheung, A., Cheng, K., & Zeil, J. (2008). The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(1), 1.

Wang, R. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., ... & Hsieh, B. B. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. *Psychonomic Bulletin & Review*, *13*(2), 281-286.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*(2), 191-210.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749-750.

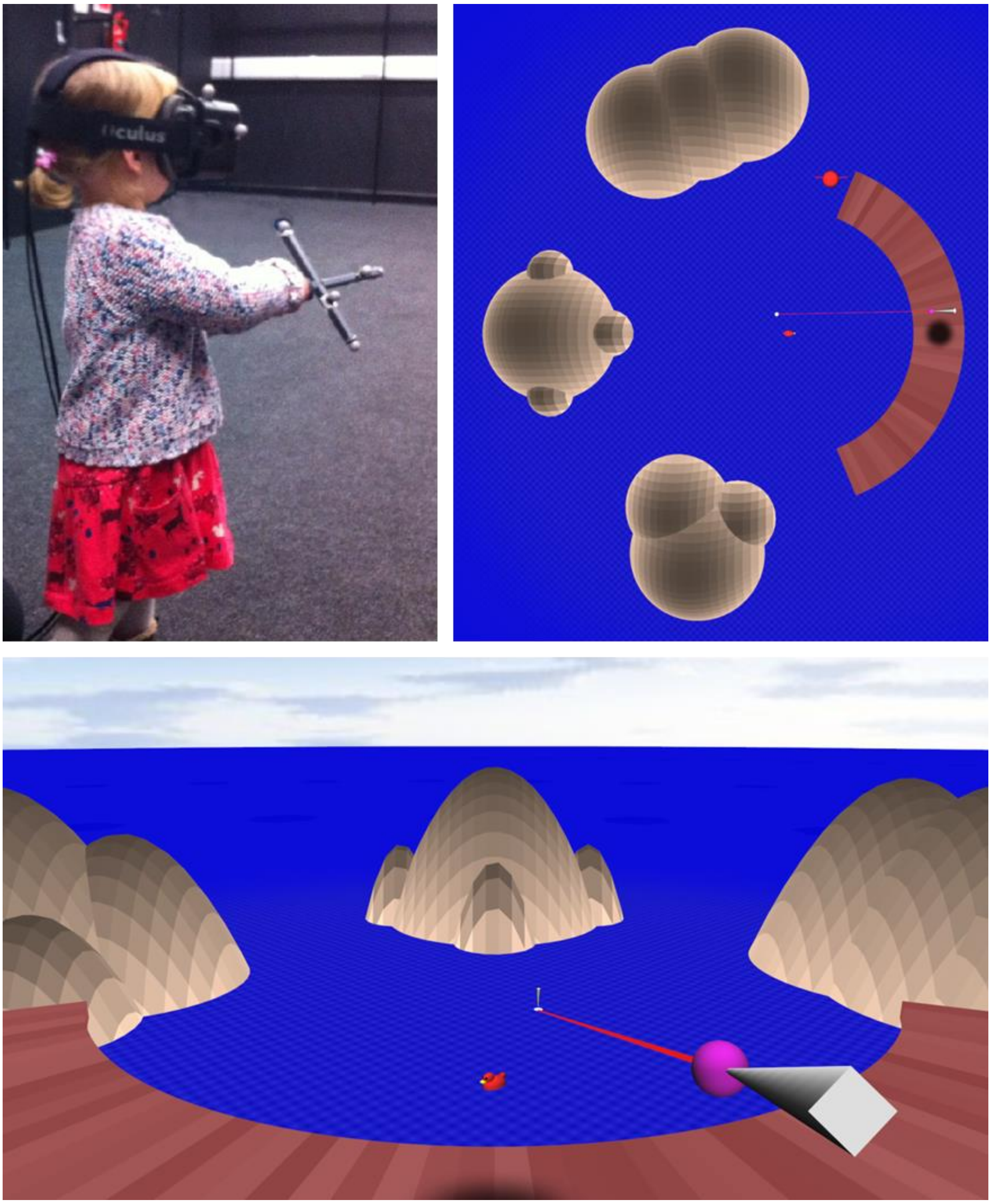

Figure 1: A three-year-old child holding the wand and wearing the display equipment (top left), an overhead screenshot of the environment (top right), and a perspective screenshot of the environment (bottom). The object that the child is holding projects onto the wand in the environment. The small red duck is the target to remember. The red sphere with wings is the experimenter's character. Movement of both the headset and the wand is displayed on a one-to-one scaling in the virtual environment.

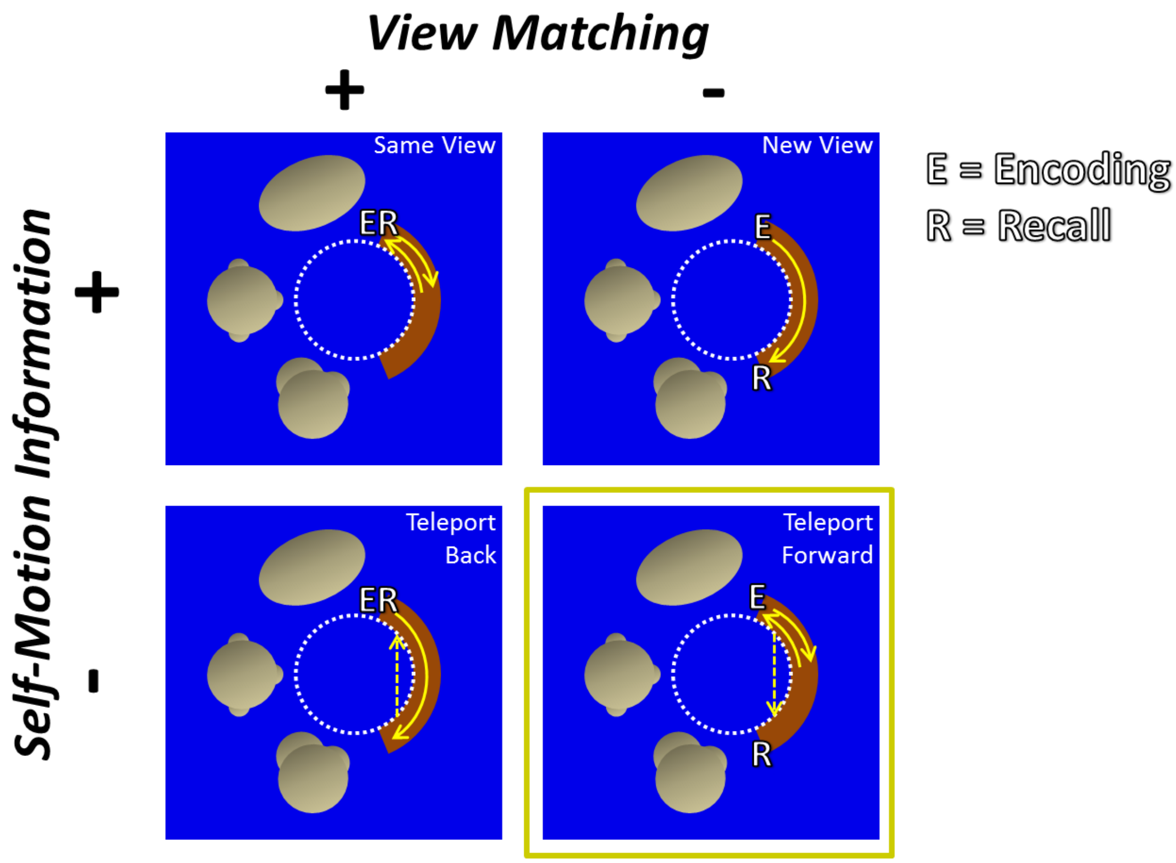

Figure 2: The different conditions from an overhead perspective. (Shown diagrammatically but to scale; see Figure 1 for real views from above and from the walkway.) The solid yellow arrows are walking paths. The broken yellow arrows are teleportation paths. The bottom right condition, *teleport forward*, is our main test of allocentric recall, since it eliminates self-motion information (via teleporting) and requires them to recall from a new viewpoint.

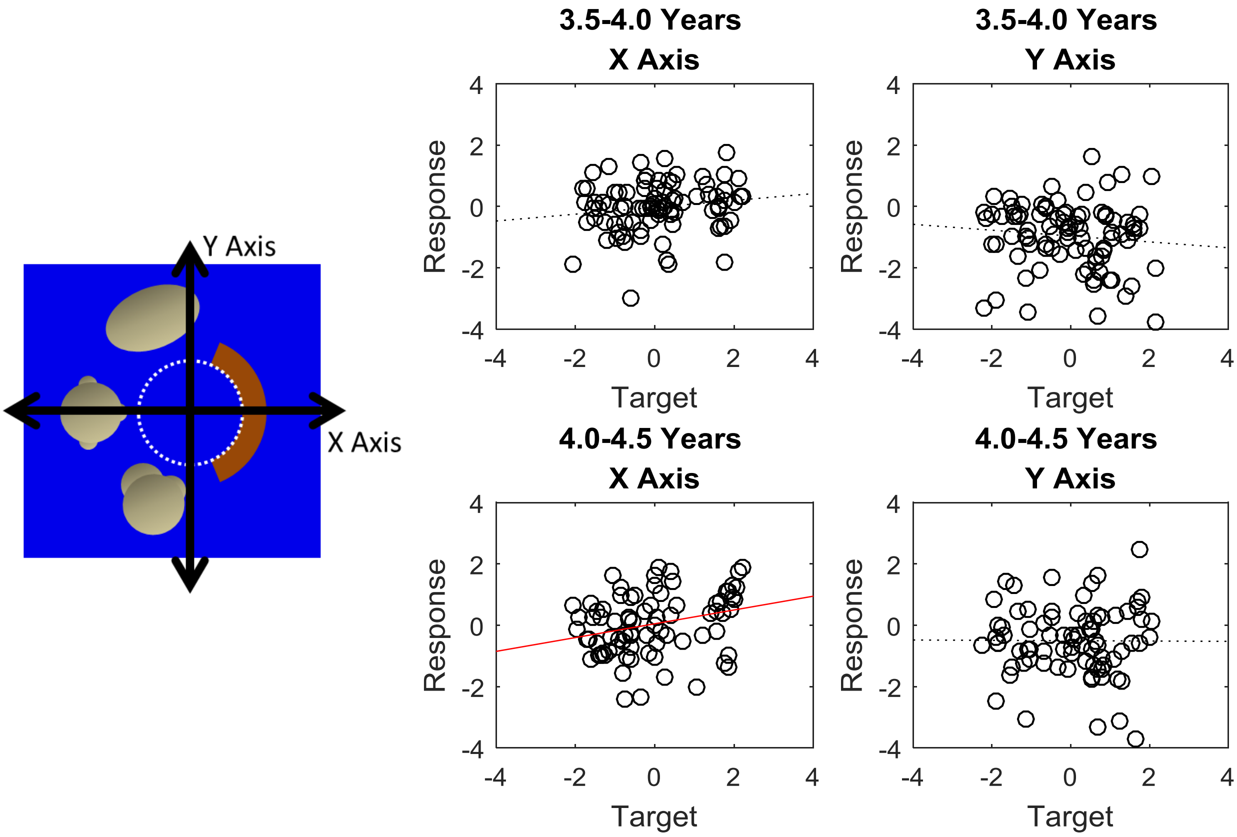

Figure 3: (Left) The X and Y axis of the task layout. (Right) Children at 4.0-4.5 years gave responses that were significantly correlated with targets along the X axis on the *teleport forward* trials, but not younger children and not along the Y axis.

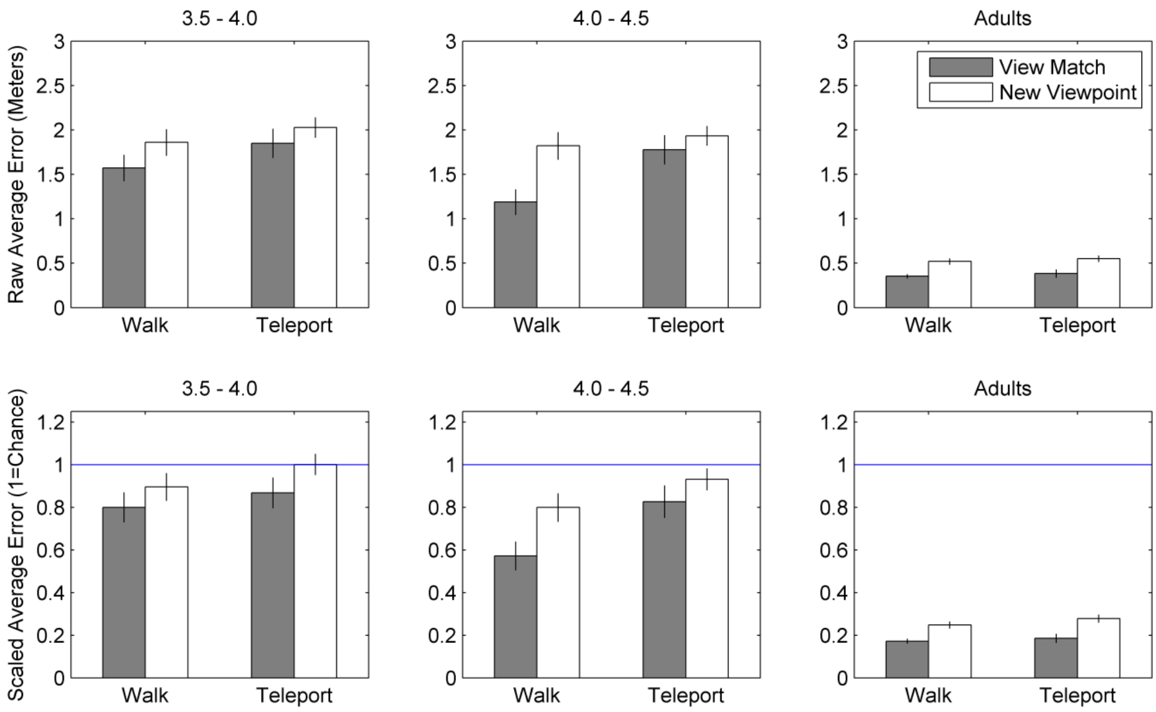

Figure 4: Average raw (top) and scaled (bottom) error broken down by age group and condition (left to right: *same view, new view, teleport back*, *teleport forward*). Error bars are standard errors.

Appendix A: Full Raw Data

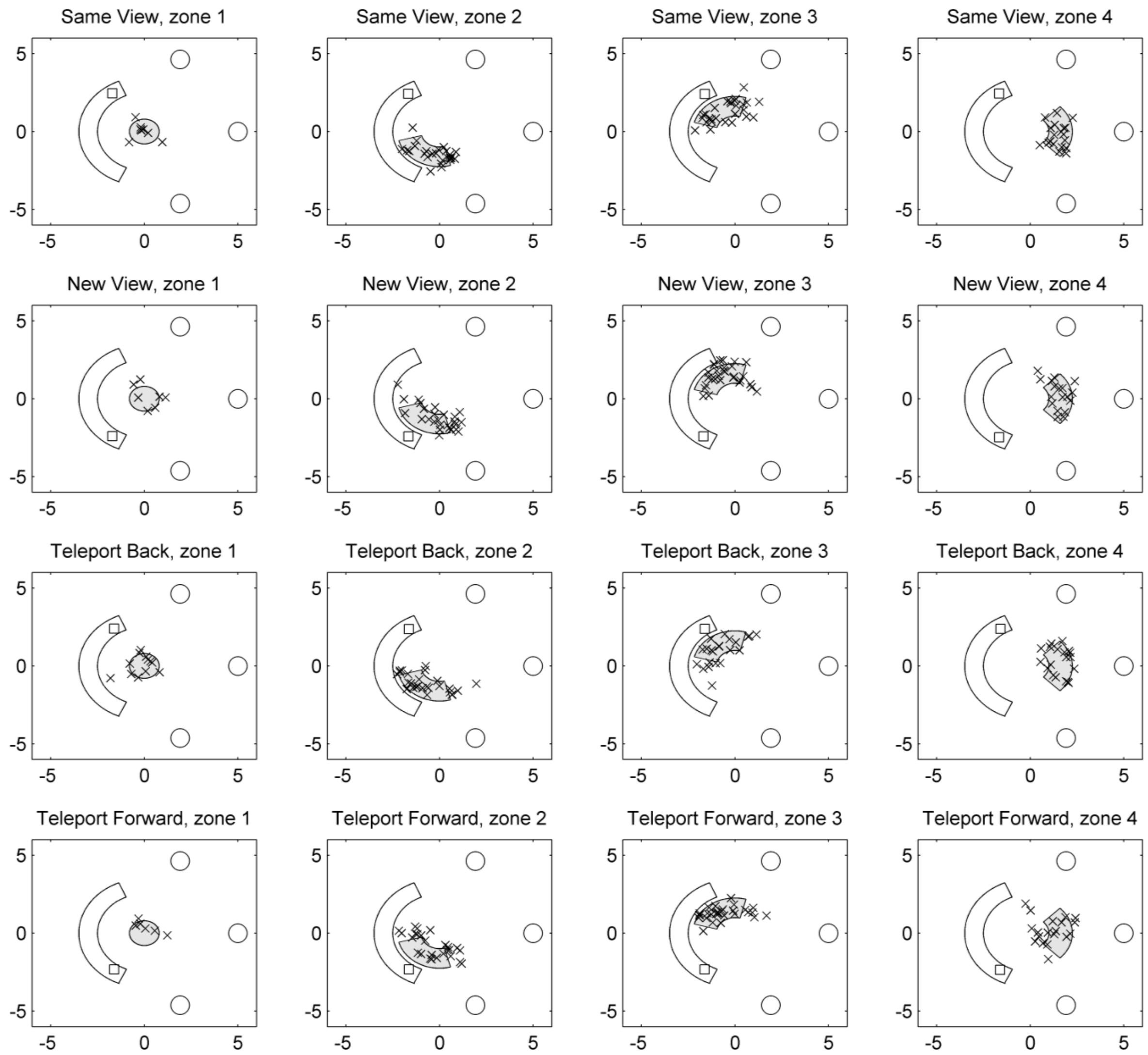

Figure A.1: Full display of the adult data. The white-filled crescent shape is the jetty and the three circles are landmarks. The shaded zone is the area where the targets appeared. Individual Xs are responses. The small square is the place where the participant stood to make the response. Axis labels are in meters. Each row is a condition. Each column is a target zone.

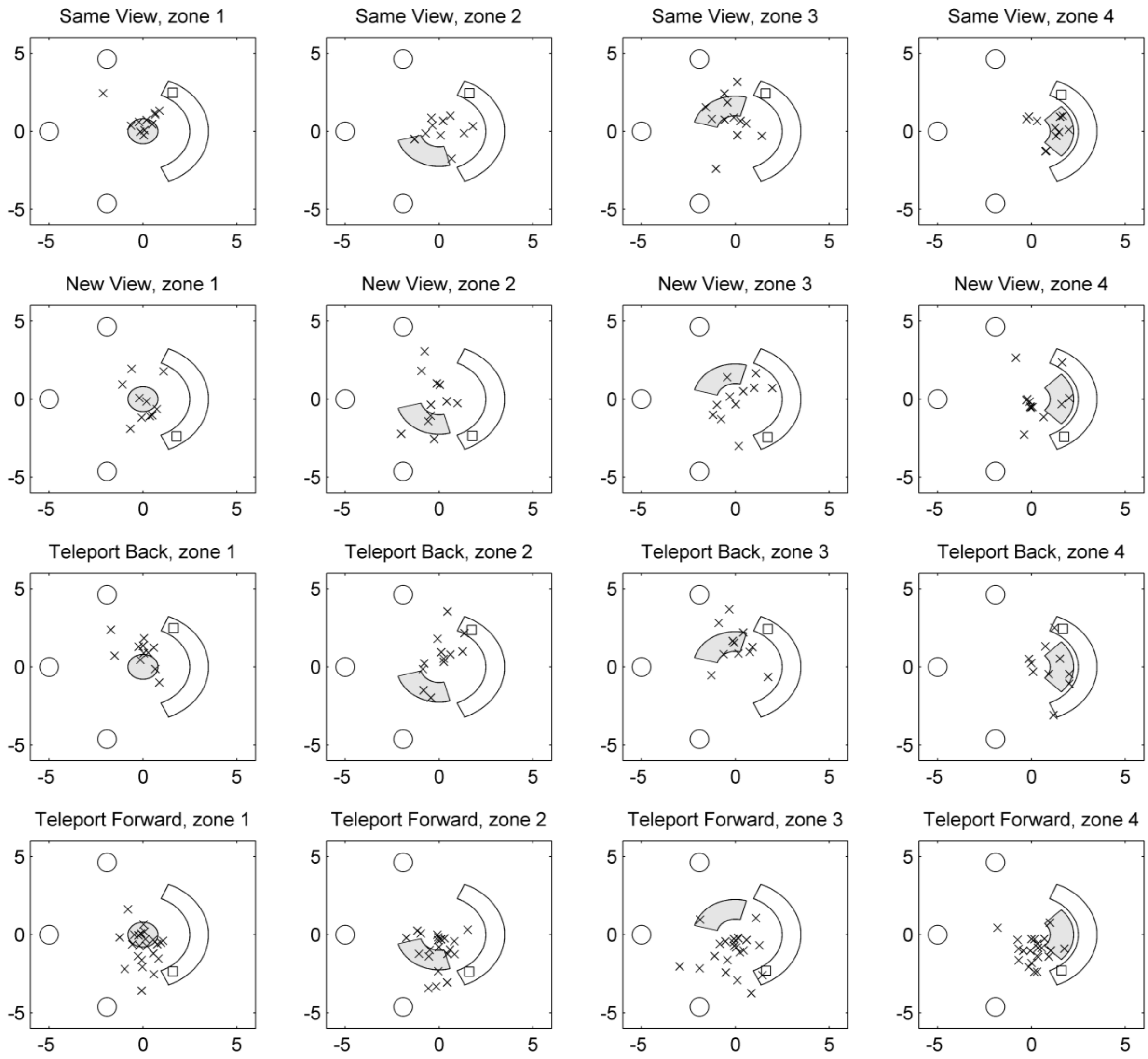

Figure A.2: Full display of the data from the younger children (3.5-4.0). When we began testing children, the virtual space was rotated 180 degrees to give watching parents a convenient place to sit without the headset wires being dragged across them.

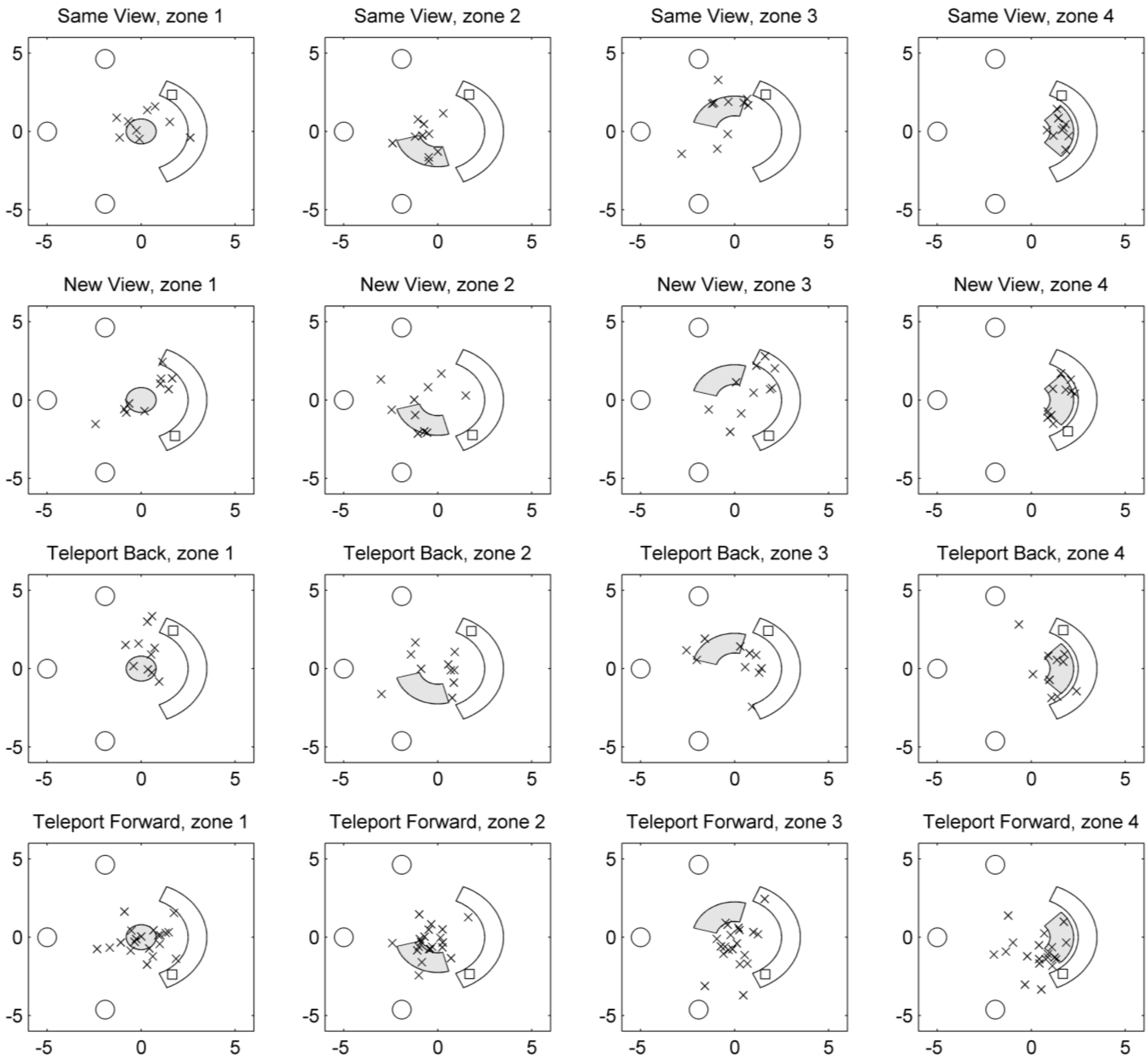

Figure A.3: Full display of the data from the older children (4.0-4.5).