***Title Page**

Brockhoff, A., Papenmeier, F., Wolf, K., Pfeiffer, T., Jahn, G., & Huff, M. (in press). Viewpoint matters: Exploring the involvement of reference frames in multiple object tracking from a developmental perspective. *Cognitive Development*.

Viewpoint Matters: Exploring the Involvement of Reference Frames in Multiple Object Tracking from a Developmental Perspective

Alisa Brockhoff[1], Frank Papenmeier[1], Kerstin Wolf[2], Till Pfeiffer[2], Georg Jahn[3], Markus Huff[1]

Authors' Note

[1]Department of Psychology, University of Tübingen, Germany;

[2] University of Education Karlsruhe

[3] University of Lübeck, Germany

Correspondence concerning this article should be addressed to Alisa Brockhoff, University of Tübingen, Department of Psychology, Schleichstr. 4, D-72076 Tübingen, Germany. E-mail: alisa.brockhoff@uni-tuebingen.de

Research Highlights

- Children of age 6 are already able to track multiple objects across viewpoint changes.
- Children use scene-based reference frames for attentive tracking.
- There were no differences in strategy used between different age groups.
- We propose that scene-based reference frames are connected to global motion procession.

Abstract

Earlier studies demonstrated that visual tracking of dynamic objects is supported by both scene-based and object-based reference frames, depending on the magnitude of scene displacement (Huff, Jahn, & Schwan, 2009; Liu et al., 2005). The current experiment tests if this pattern also applies to younger participants, i.e. school-age children, by comparing the effects of abrupt scene rotations on tracking performance of multiple dynamic objects in a 3D scene across five age groups (grade 1, 3, 5, 7 and adults). Scene rotations have two consequences: displacement of (1) the whole scene and, (2) individual objects. Tracking accuracy of 123 participants was measured across five age groups (grades 1, 3, 5, 7, and adults). Either 1 or 3 targets moved independently among a total of 8 identical objects for 5 seconds. The scene remained constant or was rotated by 10° or 20° after 3 seconds. Tracking performance of all participants was well above chance level (probability of 0.5) and an age-related increase in performance was observed. Contrasting the two factors revealed that scene rotation had a greater impact on performance than object displacement. Further, the effect of abrupt rotations was independent of age. These findings suggest that allocentric reference frames support attentive tracking across abrupt viewpoint changes and that scene-based tracking is already applied early in human development. Findings are discussed in light of new studies that link MOT to grouping processes (local and global). We propose that scene-based or allocentric processing abilities undergo a similar development as, or are connected to, grouping skills.

Processing motion of cars or other road users in heavy traffic requires several attentional skills (Dunbar, Hill & Lewis, 2001; Tabibi & Pfeffer, 2007), but mainly to keep track of multiple moving objects. Traffic is just one of many everyday life examples illustrating the importance of studying the development of tracking skills in children and young adults.

Attention allocation in complex dynamic environments is experimentally tested using the Multiple Object Tracking Paradigm (MOT; Pylyshyn & Storm, 1988). While watching several identical moving objects, observers are asked to maintain focus on a pre-assigned group of target objects. Developmental studies demonstrated that the number of objects children can track simultaneously increases markedly between 3 years of age and adulthood (Dye & Bavelier, 2010; O'Hearn, Hoffman, & Landau, 2010; Trick, Audet, & Dales, 2003;Trick, Hollinsworth, & Brodeur, 2009; Trick, Jaspers-Fayer, & Sethi, 2005). However, the majority of studies has focused on children over the age of 5, except for O'Hearn, Landau, and Hoffman (2010) who tested typically developing 3- and 4-year-olds and people with Williams Syndrome on multiple object tracking (MOT) and memory for static spatial location. Less is known about which maturing system is contributing to or is responsible for the observed improvement. O'Hearn et al. (2010) suggest that the developing visuospatial working memory (see also Klingberg, 2006) or attentional resolution (Wolf & Pfeiffer, 2014) play a role, whereas others see the number of tracked objects as reflecting the limited capacity of the maturing attentional system (Alvarez & Franconeri, 2007; Trick et al., 2005). MOT studies involving young individuals with disorders (e.g., Autism Spectrum Disorders (ASD), Williams Syndrome, Fragile X syndrome, and, Turner's syndrome) who typically showed a lower mean of successfully tracked objects (Farzin et al., 2010; Beaton et al. 2010; Hahler et al. 2010; O'Hearn et al. 2005, 2010, 2011) suggest that MOT may even be utilized as a screening tool to measure a developmental delay in different developing groups during childhood.

In addition to a developmental trend in tracking ability, attentional tracking may change qualitatively with age and experience. MOT tasks presenting objects in 3-D scenes enable the exploration of visuospatial attention during attentive tracking, with regard to the question of whether reference frames are used during MOT tasks, and if so, which ones. Humans use reference frames to transform scattered visual information input into one stable and detailed representation. When constructing a reference frame, it is possible to use objects, the environment, or the viewer as reference points (Howard, 1982). At present, there is little agreement on the form of reference used during attentive tracking. Liu et al. (2005) have speculated that MOT mechanisms in 3-D scenes only rely on allocentric, scene-based coordinates. Thus, referencing objects in relation to each other would make attentive tracking robust against abrupt viewpoint changes – i.e., the displacement of objects by cuts from one camera perspective to another should not influence tracking performance. To test this speculation, Huff, Jahn, and Schwan (2009) introduced scene rotations of 10°, 20°, and 30° to a MOT task that was adapted to 3-D. The authors hypothesized that allocentric representations are only necessary for a successful relocation of objects in cases of large viewpoint changes. Minor rotations, however, change retinocentric coordinates only minimally. Because tracking performance was significantly decreased in 20° and 30° conditions, but not for 10° rotations, they concluded that the visual system relies on the retinocentric framework and compensates for small displacements when tracking multiple moving objects. The authors attempted to test for the involvement of retinocentric processes by using the screen coordinates of objects and calculated their displacement in conditions with rotation. The extent of object displacement was analyzed for trials with 30° viewpoint changes and two targets, finding no effect between large and small displacement for targets far and close to the center of rotation, respectively. Thus, not the displacement of an object but the rotation of the whole scene determined tracking performance.

Scene-based processing presupposes the ability to integrate local sensory information into one global whole. The ability to reference objects in relation to each other, perceiving them globally as one dynamic structure, overcomes the capacity limitations of selective attention (Yantis, 1992) and makes tracking robust against abrupt viewpoint changes (Jahn, Papenmeier, Meyerhoff, & Huff, 2012). In MOT, this ability was discussed in light of the target grouping approach by Yantis (1992) who argues that tracking benefits from grouping the single targets into one higher-order object, such as three targets into a triangle. Recent studies by Evers et al. (2014) and Van der Hallen et al. (2015) modified a MOT task to explore grouping interference in normally developing children and children with ASD (autism spectrum disorder). Both research teams picked up the approach by Scholl, Pylyshyn and Feldman (2001), namely that target objects in MOT are units of attentional selection. They paired each target with a distractor by displaying a connecting line between them and compared the tracking performance to trials in which objects were left ungrouped. If the performance in the grouped condition was significantly worse than in the ungrouped condition, one can assume that global processing, i.e. perceiving objects as connected to each other, interfered with the tracking task. And in fact, global processing in MOT was measured based on a weaker tracking performance in the grouped condition, supporting the idea that grouping may shape sensory processing throughout the whole life span (Carey & Xu, 2001). Another recent study by O'Hearn et al. (2013) compared adults, children, and matched participants with autism on a modified MOT task. The multiple objects were grouped in two ways, first by arranging them, i.e. varying the space between them, to imply a grouped element and second, by letting them move together. This design allowed the authors to compare performance, for example, on target-target and target-distractor trials. They found children aged 9-12 years to show the same influence of motion-based, as well as element-based grouping as adults. Processing of the scene rather than single objects may evolve to enhance tracking performance, for example, when target objects are perceived as connected.

Scene-based, global processing has been observed in various studies using dynamic stimuli and different samples of clinical and typically developing children but it has not been explored whether this ability is under development, i.e. whether this ability partially explains the developmental curve of tracking performance in children.

Taken together, the current paper strives to answer the question whether tracking performance in children is determined only by object-based (local) processes or also by scene-based (global) processes. We assume that allocentric, scene-based processing and a global perception of multiple objects as a single grouped element are closely related, if not the same in a task in which the objects are displayed on a floor plane that is abruptly rotated in 3D (Jahn et al., 2012). The abrupt rotation of the floor plane has two consequences: the displacement of the individual objects and the rotation of the whole scene. If observers are tracking multiple objects in an object-based local manner, only the displacement of each individual object should determine tracking performance (lower tracking performance the further an object is displaced). However, if observers also utilize scene-based information such as grouping multiple objects into a higher-order object, the amount of scene change (angle of abrupt scene rotation) should explain tracking performance over and above the displacement of individual objects alone. Based on what we know about the effect of grouping in MOT studies (e.g., Van der Hallen et al., 2015), we expected to find scene-based effects across all age groups tested.

To shed light on how attentionally-demanding visuospatial skills mature with age, a more detailed analysis will focus on the strength of each impact on different age levels. To our knowledge, this is the first study that tested different age groups to see whether scene rotations impair tracking performance less with increasing age. Because adults are more experienced in global processing, an alternative finding would be that adults' tracking performance is even more impacted by scene rotations than children's performance.

Method

Participants

The sample consisted of 123 participants. Twenty-seven children were in grade 1 (age in years: *M*=6.45, *SD*=0.57), 31 in grade 3 (*M*=8.71, *SD*=0.55), 23 in grade 5 (*M*=11.51, *SD*=0.49) and 23 in grade 7 (*M*=13.34, *SD*=0.50). In sum: 104 children completed the experiment at the University of Education in Karlsruhe after written consent was obtained from parents. Seventeen adults participated (15 from the University of Education in Karlsruhe and 2 from the University of Tübingen). Three participants were excluded due to technical issues during the experimental session. The children received a small present for their participation and the adults were given monetary compensation.

Stimuli and procedure

Stimuli were presented using the Blender game engine (www.blender.org) and custom software written in Python. They were 8 small white, black-bordered 3-D spheres moving on a checkerboard floor plane (see Figure 1). At the beginning of each trial, the 8 spheres were randomly positioned on screen. After 2 s, 1 or 3 spheres flashed red 4 times within 1.6 s and remained red for another 2 s. These spheres were the target objects. The target spheres turned white again and all spheres began to move at a constant speed of 3°/s for 5 s. The spheres moved in random directions and were allowed to touch or to overlap. Reaching the boundaries of the checkerboard, the spheres were reflected in a physically consistent manner (comparable to billiard balls), however the spheres did not bounce off of each other.

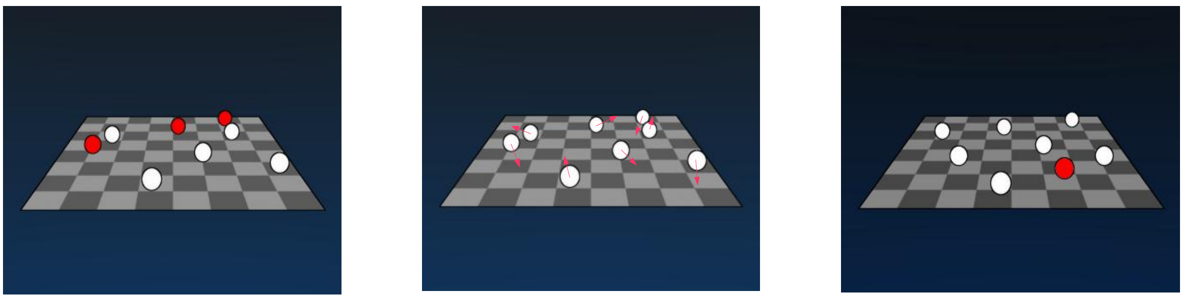

Figure 1. Target designation, visual tracking and decision/marking phase.

The rotation of the scene was characterized by 3 conditions: the scene either remained constant, or it was rotated by 10°, or 20° (around the vertical axis through the center of the floor rectangle). It appeared abruptly (as if a camera cut in a movie displayed the same scene from another person's view) and did not influence the movement of the spheres. Rotations occurred after 3 s. Half of the rotations were directed to the left, the other half to the right. Figure 2 illustrates a simplified rotation to the right and the two emerging variables we used for the analysis (see the next section for more details).

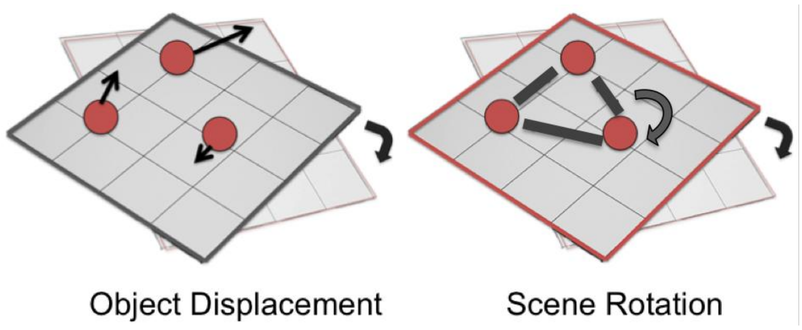

Figure 2. A simplified visualization of the moment of rotation with different foci: perceiving objects as individually displaced (left) or as a group of objects rotated (right).

Following randomized movement, the spheres came to a stop and one turned red. The observer, then, had to indicate whether the marked object was part of the original target set seen at the beginning of the trial. Demo videos can be found here: https://homepages.uni-tuebingen.de/frank.papenmeier/mot-develop/.

Participants proceeded to the next trial by pressing the spacebar. Each participant performed 6 practice trials (2 levels of target number x 3 levels of scene rotation). The final experiment was comprised of 72 trials (2 target numbers (1 or 3) x 3 levels of scene rotation (0°, 10°, or 20°) x 12 repetitions). The order of conditions was randomized throughout the experiment. The participants had the option to take self-paced breaks between the trials. The within-subjects design allowed for controlling individual differences, reducing the associated error variance. The different grades (1, 3, 5, 7, and adults) served as a between-factor.

Data Analyses

The application of a mixed-factor ANOVA on the proportion of correct answers provided a first impression of the data. In a further analysis, we fit logistic generalized mixed-models (*glmer*) due to the non-linear response variable that was expressed as a categorical variable with two levels (Yes/No). The aim was to quantify age-related and inter-individual differences that might influence the factors *scene rotation* and *object displacement* that, in turn, were thought to determine the variability in the number of correct responses. *Object displacement* was calculated as the distance on the screen between the location of the target probe right before and right after the rotation. We constructed *object displacement* as a continuous factor. Because a 0° scene rotation would automatically result in a displacement of 0 degrees of visual angle, only 10° and 20° trials were analyzed within *the glmer* analysis.

The *lme4* package for R (Bates, Maechler, Bolker & Walker, 2014) was used to perform the binomial logistic analysis. In a first step, using likelihood-ratio tests, the fit of the model with only object displacement as a fixed effect was compared to the fit of a model including both scene rotation and object displacement as fixed effects, in order to investigate whether scene rotation has a beneficial contribution. We called this the “Scene over Objects” logic: A significant result would lead to the acceptance of the model with both effects. Thus, both scene rotation and object displacement would contribute to successful tracking. A non-significant result would lead to the rejection of the model with both effects, i.e. indicate that object displacement explains the variance sufficiently. In a second step, we compared the fit of the model with only scene rotation to the fit of the model including both. We tested whether object displacement has an additional explanatory benefit. This “Objects over Scene” logic is similar to the “Scene over Objects” logic with the order of including the fixed effects into the models interchanged. Participants, specified as a random effect, allowed a separation of between-subjects (inter-individual) and within-subjects (responses to the variable of interest depending on individual differences) variance in the data.

Results

Repeated-Measures Analysis of Variance

The mixed factor ANOVA has been executed with the between-subject factor *age* (grades 1, 3, 5, 7 and adults) and the within-subjects factors *number of targets* (1 or 3) and *level of scene rotation* (0°, 10°, and 20°) on the mean proportion of correctly identified targets. As predicted from previous research, statistically significant main effects of age, $F(4, 116) = 16.35$, $p < .001$, scene rotation $F(4, 116) = 23.87$, $p < .001$, and number of targets $F(4, 116) = 138.94$, $p < .001$ on mean proportion correct were observed. The effect of level of scene rotation was the same for all age groups, $F(8, 232) = 0.55$, $p = 0.82$, whereas age and number of targets as well as scene rotation and number of targets appeared to interact, $F(4, 116) = 4.58$, $p < .001$; $F(2, 232) = 4.79$, $p < .001$, respectively (see Figure 2).
Based on established findings in the literature it is not surprising that an increased tracking load decreased performance in young participants. Further, the influence of the number of targets was higher in conditions with larger scene rotations. Finally, the interaction of age, scene rotation, and number of targets was not significant, $F(8, 232) = 0.73$, $p = 0.67$.

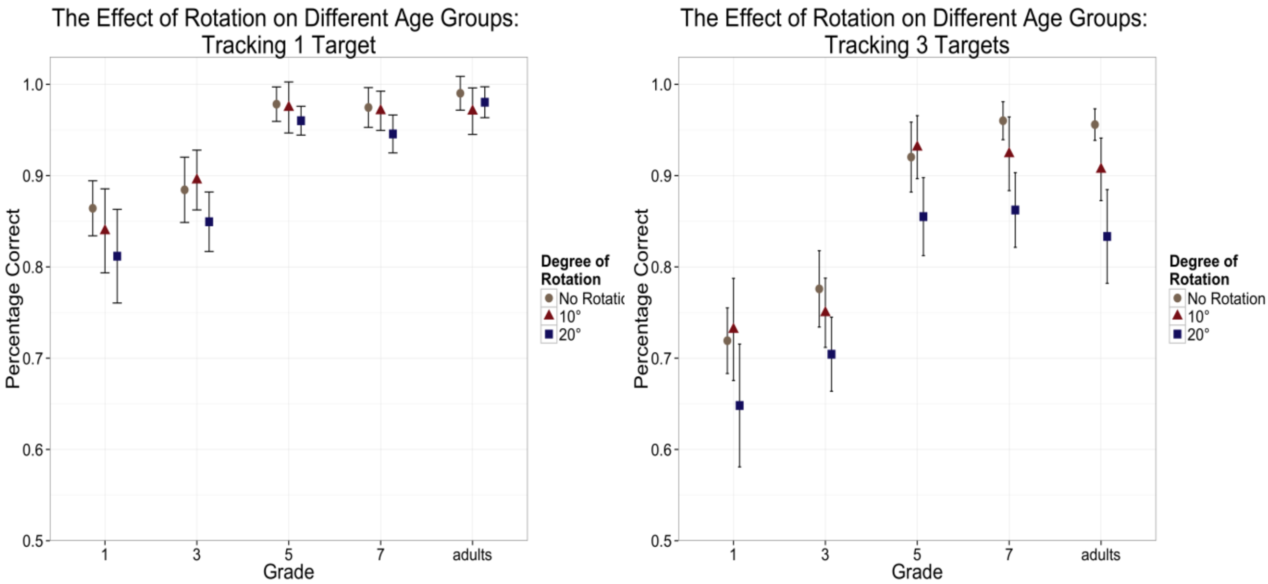

Figure 3. Influence of age and number of targets on proportion correct, separated by scene rotation. Error bars indicate the 95% confidence interval of the mean.

Generalized mixed-effects models for object displacement and scene rotation

The repeated measures ANOVA provided a first impression of the data, suggesting an exponential, developmental nature of tracking skills, with scene rotation having the same effect for all age groups. In a further analysis, we explored which reference frame (allocentric or retinocentric) participants used across the age groups. Therefore, we ran a separate analysis on all target probe trials with scene rotations of 10° and 20° and calculated the object displacement of the target probe. The direction of the data (see Figure 3) mirrors the predicted, developmental trajectory nicely, but also points to a lower performance in trials with 20° rotation independent of object displacement. To better understand the different effects of the two predictors (scene rotation and object displacement), tracking performance (correctness of response) was subjected to a binomial *glmer* with the factors scene rotation (10° or 20°) and object displacement (M = 24.53 pixels, range [0.00; 125.40] pixels).

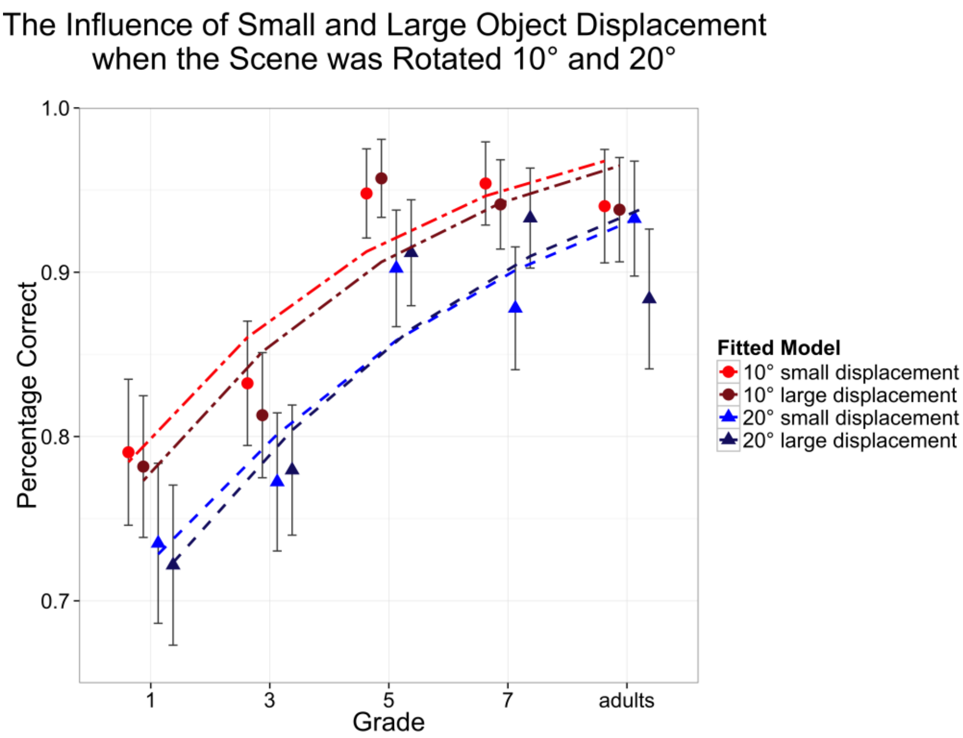

Figure 4. Influence of object displacement and rotation on tracking performance over all target-trials. The red line and blue lines depict 10° and 20° rotation, respectively.

For better visualization only, we utilized the median as a cut-off score and displayed *object displacement* with two levels (small/large). Note that the factor was continuous in the analysis. Error bars represent 95% confidence intervals.

Based on previous research results by Huff et al. (2009), we surmised that object displacements are means to study the use of retinocentric reference frames. We assume that global and local processing, i.e. the extent of perceiving the objects as a group, will determine how much influence the rotation of the scene or the displacement of target objects has on tracking performance. To this end, we applied the “scene-over-objects“ and the “objects-over-scene” logic for each grade and the adults separately. A side-by-side comparison of the results of each model by grade, as well as exact *p*-values, can be found in Table 1. Applying the scene-over-objects logic resulted in the acceptance of the model including scene rotation *and* object displacement. Thus, object displacement alone is not sufficient to explain the variance. In a second step, we applied the objects-over-scene logic. Results showed that the model with only scene rotation provided the best fit for grade 1 ($\chi^2(1) = 5.20$, $p = .023$), grade 5 ($\chi^2(1) = 4.77$, $p = .030$), grade 7 ($\chi^2(1) = 9.17$, $p = .003$) and adults ($\chi^2(1) = 13.87$, $p < .001$). Surprisingly, for grade 3, the scene-over-objects logic accepted the model with object displacement as fixed effect and the objects-over-scene logic accepted the model with scene rotation as fixed effect ($\chi^2(1) = 0.49$, $p = .288$). These findings stand in contrast to those of all other age groups. Thorough analysis neither revealed extreme outliers, nor an increased rate of guessing (calculated as proportion correct smaller than 0.5), or misunderstanding of the task (measured as participants pressing only one key, i.e. saying “Yes” or “No” constantly). Therefore, and based on the consistent picture of all other grades and adults demonstrated, we can only assume these effects to be due to random variation. Taken together, scene rotation was not only integrated into the tracking task but was a significant predictor of performance.

Table 1. *Generalized mixed-effects models for object displacement and scene rotation*.

	Scene-over-Objects		Objects-over-Scene	
Grade	χ^2	*p*	χ^2	*p*
1	5.20	.023*	0.09	.783
3	0.49	.288	1.19	.503
5	4.77	.030*	0.38	.532
7	9.17	.003*	0.002	.959
Adults	8.97	.003*	1.89	.163
Reanalysis of Huff et al. (2009)	13.87	<.001*	0.36	.549

For the sake of completeness, we reanalyzed the original data published by Huff et al. (2009). The displacement range of an object and viewpoint change played a part in predicting performance in the original data set as well. Put in contrast by using the scene-over-objects/ objects-over-scene logic, we found again, that the scene was superior over displacement of objects in predicting performance ($\chi^2(1) = 13.87$, $p < .001$), even when controlling for individual differences and varying speed – providing further support for the importance of the scene over an object during tracking.

Discussion

It has been the subject of considerable debate which representations visuospatial attention accesses during tracking processes (Huff et al., 2009; Seiffert, 2005; Liu et al., 2005). Huff et al. (2009) left allocentric coordinates intact and still found an impaired tracking performance. Although this points towards a retinocentric, viewer-based representation of dynamic scenes, other interpretations are possible. The focus of the current study was to replicate preceding results of studies concerning the usage of reference frames during tracking – concentrating in particular on the development of tracking abilities in younger participants in 3-D environments. The results presented here indicate that the impact of rotations is similar across all age groups tested – independent of the range of object displacements. These findings are in line with recent studies that linked global processing of objects to MOT as well. Evers et al (2014) suggested a reduced global processing bias in participants with ASD compared to normally developing children (see also O'Hearn et al. 2013). Grouping of targets and distractors (paired by a connecting line) resulted in an interference of the tracking task, suggesting that forming object-based connections (grouping) is a tracking approach observable already in young children. If children and adults track multiple objects by utilizing scene-based processes such as grouping target objects to a higher-order object (e.g., a triangle), tracking across abrupt scene rotations should not only be influenced by the displacement of individual objects caused by the rotations but also by the extent of the scene rotations as such.

Does sole rotation of the whole scene or the extent of the displacement of a target object influence tracking performance?

The first part of the analysis addressed the question of which factors affect tracking performance. We found our results to replicate established findings in MOT research with main effects for number of targets, age, object displacement, and scene rotation. But which factor produces more tracking errors? For further examination, we introduced a new way of modeling tracking performance in relation to object displacement and scene rotation. Interestingly, scene rotation was a better predictor of tracking performance than object displacement (objects-over-scene logic). The finding that scene rotation influenced tracking performance more than object displacement leads us to speculate that humans not only rely on retinocentric changes, but also make use of scene-based, allocentric reference frames, especially during tracking tasks in 3-D environments.

Are viewer-based effects also observable in younger participants – and given that both rotation and displacement reveal an impact, which one is stronger?

All groups showed a similar pattern of performance drop due to object displacement and scene rotation, which may be continuous throughout development. Assuming that tracking processes are retinocentric in nature, larger object displacement should result in a higher number of errors. The current results only partially support this assumption. Scene rotation was a better predictor of performance than object displacements, suggesting a strong involvement of allocentric processes during tracking in 3-D environments. These results rather coincide with speculations by Liu et al. (2005) who surmised that a critical input needed for tracking multiple objects is a stable environment, not the objects themselves. The superior influence of scene rotations was present in almost all grades and conditions tested, leading to the conclusion that scene-based, not viewer-based effects are observable already in children of age 6.

Future Research

Future research may determine (a) the extent of developmental effects on the use of reference frames in tracking and (b), the connection of reference frames and grouping. A field to apply this knowledge could be the design of perceptual-cognitive training games in dynamic, virtual reality environments, to help improve tracking-speed and tracking-capacity in order to reduce the risk of road accidents for children. It has been shown that age-related effects in tracking can be reduced by training for older participants (e.g., Legault, Allard & Faubert, 2013). Intelligently designed dynamic environments may be used to teach attentive tracking and visuospatial skills.

Logan (1995) included linguistics in tasks of spatial representations and proposed linguistic cues to play a role in directing attention. Trick, Audet, and Dales (2003) suggested a relationship between tracking and enumeration. This could explain the considerably large difference found between primary school children and grades 5 and 7: language and enumeration skills, as well as tracking, all undergo huge improvement between childhood and young adulthood. Further research exploring these skill combinations in depth will be interesting with regard to the development of underlying cognitive skills and reference frames needed in tracking tasks.

But not only the application of reference frames should be specified in more detail. Generally, children are assumed to be less efficient in their deployment of attention (e.g., Plude, Enns, & Brodeur, 1994; Trick & Enns, 1998) whereas adults can make use of more than one reference frame simultaneously (Carlson-Radvansky & Jiang, 1998; Stein, 1992). Thus, the gradual improvement of sustained attention and longer periods of extended concentration may play a role, as well as visual working memory and attentional selection of items.

The most reasonable approach based on current literature and our recent results would be to assume that allocentric and retinocentric frames (or global and local processes) are at work simultaneously. It is possible that people develop a strategy to track objects in a global or local manner, possibly by activating different reference frames or using processing strategies that are applied depending on the situation. Whether multiple intrinsic representations are accessed in a top down manner, as well as when and if a strategy develops and why, this has yet to be determined.

Conclusion

By exploring developmental processes, we were able to show that the magnitude of age-related changes is consistent over different ages and depends on the stimulus complexity (number of targets, range of displacement, and extent of rotation). Concerning the hypotheses, the results indicated that (1) object-based effects are observable from early age on, but are less pronounced than scene-based effects, (2) scene rotation and displacement of targeted objects have an influence on tracking performance, and finally, we showed that (3) scene rotation had a stronger impact than object displacement, leading us to assume that tracking across abrupt viewpoint changes in 3-D environments relies more on allocentric than on retionocentric processes.

The findings of the presented experiment offer numerous theoretical and practical implications. Within the context of perceptual developmental theories on grouping processes, our measure of children's performance in situations of scene or object shifts brings us closer to understanding how attentionally-demanding visual-spatial skills mature with age. The limited tracking ability of children in grade 1, 3, and 5, relative to adults supports existing, findings suggesting that brain areas responsible for MOT develop and only become maximally efficient later in life (see Ryokai et al, 2013; Dye & Bavelier, 2010; Trick et al. 2003). The similar influence of scene rotation on all groups suggests that grouping, i.e. processing the presented objects in a global manner, is already present in children as young as 6 years. By documenting a specific window of time of the typical developmental trajectory of the use of reference frames during tracking, we can learn more about how children experience and structure their complex environments. Our results, and maybe even our version of the MOT task that was designed in a game-like manner, may guide parents, teachers, clinicians, and researchers in identifying developmental delays in scene-based motion processing.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6.http://CRAN.Rproject.org/package=lme4

Beaton, E. A., Stoddard, J., Lai, S., Lackey, J., Shi, J., Ross, J. L., & Simon, T. J. (2010). Atypical functional brain activation during a multiple object tracking task in girls with Turner syndrome: neurocorrelates of reduced spatiotemporal resolution. *Journal Information*, *115*(2).

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, *80*(1), 179-213.

Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference-frame selection. *Psychological Science, 9*(5), 386-391.

Dunbar, G., Hill, R., & Lewis, V. (2001). Children's attentional skills and road behavior. *Journal of Experimental Psychology: Applied, 7*(3), 227.

Dye, M. W., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision research, 50*(4), 452-459.

Evers, K., de-Wit, L., Van der Hallen, R., Haesen, B., Steyaert, J., Noens, I., & Wagemans, J. (2014). Brief report: Reduced grouping interference in children with ASD: Evidence from a multiple object tracking task. *Journal of autism and developmental disorders*, *44*(7), 1779-1787.

Farzin, F., Rivera, S. M., & Whitney, D. (2010). Spatial resolution of conscious visual perception in infants. *Psychological science*.

Howard, I. P. (1982). *Human visual orientation*. Chichester: Wiley.

Huff, M., Jahn, G., & Schwan, S. (2009). Tracking multiple objects across abrupt viewpoint changes. *Visual Cognition, 17*(3), 297-306.

Jahn, G., Papenmeier, F., Meyerhoff, H. S., Huff, M. (2012). Spatial reference in multiple object tracking. *Experimental Psychology, 59*(3), 163-173. doi:10.1027/1618-3169/a000139

Klingberg, T. (2006). Development of a superior frontal–intraparietal network for visuo-spatial working memory. *Neuropsychologia*, *44*(11), 2171-2177.

Legault, I., Allard, R., and Faubert, J. (2013). Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking. *Frontiers in Psychol*y, *4*(323). DOI: 10.3389/fpsyg.2013.00323

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance, 31*(2), 235.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*(2), 103-174.

O'Hearn, K., Franconeri, S., Wright, C., Minshew, N., & Luna, B. (2013). The development of individuation in autism. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 494–509. doi:10.1037/a0029400.

O'Hearn, K., Hoffman, J. E., & Landau, B. (2010). Developmental profiles for multiple object tracking and spatial memory: Typically developing preschoolers and people with Williams syndrome. *Developmental Science, 13*(3), 430-440.

O'Hearn, K., Hoffman, J. E., & Landau, B. (2011). Small subitizing range in people with Williams syndrome. *Visual cognition*, *19*(3), 289-312.

O'Hearn, K., Landau, B., & Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological Science, 16*(11), 905-912.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta psychologica, 86*(2), 227-272.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision, 3*(3), 179-197.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*(1), 159-177.

Seiffert, A. E. (2005). Attentional tracking across display translations. *Journal of Vision, 5*(8), 643-643.

Stein, J. F. (1992). The representation of egocentric space in the posterior parietal cortex. *Behavioral and Brain Sciences, 15*(04), 691-700.

Tabibi, Z., & Pfeffer, K. (2007). Finding a safe place to cross the road: the effect of distractors and the role of attention in children's identification of safe and dangerous road-crossing sites. *Infant and Child Development, 16*(2), 193-206.

Trick, L. M., Audet, D., & Dales, L. (2003). Age differences in enumerating things that move: Implications for the development of multiple-object tracking. *Memory & Cognition, 31*(8), 1229-1237.

Trick, L. M., & Enns, J. T. (1998). Lifespan changes in attention: The visual search task. *Cognitive Development, 13*(3), 369-386.

Trick, L. M., Hollinsworth, H., & Brodeur, D. A. (2009). Multiple-object tracking across the lifespan: Do different factors contribute to diminished performance in different age groups. *Computation, cognition, and Pylyshyn*, 79-99.

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "Catch the Spies" task. *Cognitive Development, 20*(3), 373-387.

Van der Hallen, R., Evers, K., de-Wit, L., Steyaert, J., Noens, I., & Wagemans, J. (2015). Multiple Object Tracking Reveals Object-Based Grouping Interference in Children with ASD. *Journal of autism and developmental disorders*, 1-9.

Wolf, K. & Pfeiffer, T. (2014). The development of attentional resolution. *Cognitive Development, 29,* 62–80.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive psychology*, *24*(3), 295-340.

Acknowledgement

This research was supported by German Research Foundation (DFG) Grants HU 1510/4-1, HU 1510/4-3, JA 1761/5-1, and JA 1761/5-3.

Author Contributions

Idea and research design: MH, GJ, TP, AB

Operationalization: MH, FP, GJ, KW

Data analysis: AB, FP, MH

Writing: AB drafted the manuscript and all authors provided critical revisions.