Adaptive significance of frontal eyes: Binocular disparity improves the capacity of vision

Henry Railo* [1,2], Joni Saastamoinen[1], Sipi Kylmälä[1], and Aapo Peltola[3]

[1] *Department of Psychology, University of Turku, 20014, Finland*

[2] *Turku Brain and Mind Centre, University of Turku, 20014, Finland*

[3] *Turku Game Lab, Turku University of Applied Sciences, 20520, Finland*

Word count:

Introduction and Discussion: 2363

Methods and Results: 2090

Total: 2363+2090=4453

* Corresponding author:
Henry Railo,
Department of Psychology,
University of Turku,
Finland
Email: hmrail@utu.fi
Phone: +358-45-6765227

Frontal eyes are one of the defining features of primates. The horizontal separation of the eyes enables accurate calculation of depth which greatly improves eye-hand coordination – this is considered the key factor driving the evolution of frontal eyes. Although binocular disparity also results in a tangible subjective experience of three-dimensional world, it remains unknown whether disparity also augments objective perceptual performance. By utilizing realistic immersive virtual reality stimuli, we show that binocular disparity enhances the capacity of visual perception in humans: when visual disparity was present, participants were able to individuate more objects from the background more accurately, and do this faster. Interestingly, this effect emerged without a concurrent enhancement in subjectively experienced depth. We propose that the perceptual advantage provided by binocular disparity has been important in enabling more efficient visual behavior and may have helped to drive the expansion of the brain areas that process visual information during evolution.

Humans and other primates have two forward facing eyes and a large overlap between their visual fields (Heesy 2009). One of the major benefits of this is stereopsis: integration of the slightly different viewpoints of the two eyes (binocular disparity) results in tangible subjective sensation of three-dimensional visual space, and accurate coding of depth (Barry 2009; Vishwanath 2014). While binocular disparity is known to improve eye-hand coordination (Melmoth et al. 2009; O'Connor et al. 2010; Fielder and Moseley 1996), the *perceptual* consequences of disparity have received little attention. Does the subjective experience of depth imply higher perceptual capacity? Does it improve visual discrimination, or speed-up visual reactions? This question is particularly interesting because the subjective and functional consequences of disparity have been dissociated: Subjective sensation of depth can emerge without binocular disparity (Vishwanath and Hibbard 2013), and disparity can influence behavior in the absence of subjective sensation of depth (Masson, Busettini, and Miles 1997; Cumming and Parker 1997).

Stereopsis is one of the defining features of primates (Ross and Martin 2006), and linked to the expansion of brain structures that process visual information (Barton 2004). What benefits drove the emergence of frontal eyes during primate evolution? Early theories explained the emergence of frontal eyes as adaptations to arboreal locomotion because movement in arboreal environment (for instance, when leaping between tree branches) requires accurate judgement of distances (Jones 1917; Collins 1921; Smith 1924). The arboreal theory of stereopsis has been challenged as, for instance, many non-primate species (su ch as the squirrel) are capable of skilled arboreal locomotion despite the lack of a binocular visual field (Cartmill 1974) suggesting that arboreal locomotion alone does not explain the evolution of frontal eyes. According to the alternative visual predation hypothesis stereopsis is an adaptation that enables animals to better visually locate and seize prey without moving (i.e. without motion parallax; (Cartmill 1974; Cartmill 1992)). Similarly, the camouflage-breaking hypothesis of stereopsis argues that stereopsis enables animals to detect camouflaged prey (Julesz 1971).The visual predation theory was later elaborated to concern especially nocturnal animals as a large binocular visual field meant that the animals could better detect objects in dark environments (Pettigrew 1986). However, the visual predation theory is undermined by the fact that some excellent predators do not have frontal eyes, and early nocturnal primates may have detected their prey (mainly crawling insects) using olfaction or hearing (Sussman 1991; Sussman

2013; Crompton 1995). Moreover, these smaller primates were largely omnivorous, eating lots of plant material in addition to insects, which lead Sussman ((Sussman 1991), p. 219) to argue "that the explanation [for orbital convergence] will be found in adaptations providing the fine discrimination needed to exploit the small food items available on the newly diversifying flowering plants".

As noted by Crompton (Crompton 1995), different theories of stereopsis have in common the requirement to accurately discriminate objects form the background; it seems to help animals to better resolve targets (prey, branch etc.) as discrete entities. Similarly, Dal Mutto et al. (2011) noted that in machine stereo vision systems, one of the major uses for stereo vision algorithms is to aid in scene segmentation, in addition to establishing accurate 3D geometry. Our aim here is to experimentally test whether binocular disparity indeed enhances the capacity of visual perception. Establishing this possible functional advantage provided by visual disparity is crucial to understand what factors may have driven the evolution of frontal eyes.

Random dot stereograms illustrate how stereopsis may enhance perception and enable the detection of otherwise hidden objects (Julesz 1971), but they represent extremely simplified cases: all other depth cues (than binocular disparity) have been removed, and the object is placed in front of a flat background surface. In real life the objects to be detected are embedded in scenes that include various depth cues, and the scenes are often cluttered with other objects that occupy different depths. Hence, it is important to test the question with stimuli that closely resemble natural environments. There is previous anecdotal evidence that stereopsis helps observers to detect objects in realistic scenes as aerial stereographic photographs (often with greatly exaggerated visual disparity) have been used to detect camouflaged objects since the first world war (Saint-Amour 2003). However, the influence of disparity in perceiving realistic stimuli has not been previously studied experimentally. McKee et al. (1997) has shown that visual disparity aids in detecting a moving target dot that is flanked by randomly moving distractor dots that occupy a different plane in depth. However, when a moving target is presented in the midst of randomly moving distractors, disparity provides negligible help (McKee et al. 1997). Visual masking is reduced when the mask is at different depth plane than the target, suggesting that binocular disparity provides useful information for image segregation (Wardle et al. 2010).

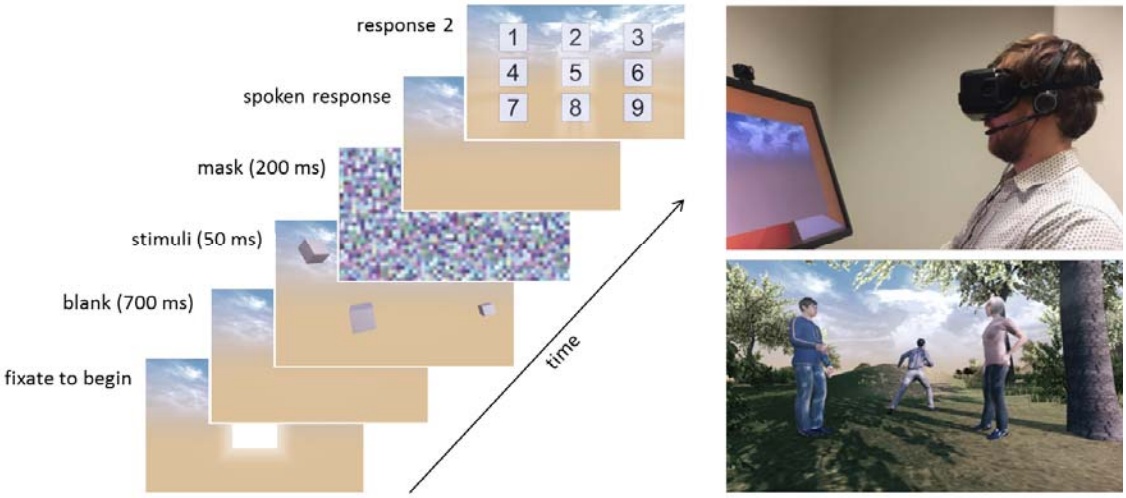

Figure 1. Task and stimuli. Schematic presentation of a single trial with cube stimuli (left; Experiment 2), the experimental set up (top right), and an example of a scene stimulus (Experiment 1). In the experiments, the stimuli (1-6 objects) were always presented binocularly and disparity was manipulated. In one condition the participants were asked to enumerate the individuals in the scene and in the other condition the participants rated the strength of their experience of depth.

We asked participants to rapidly detect and enumerate individuals presented in an immersive virtual environment. Enumeration reveals well-documented limits in both speed and capacity of segregating figures from the background: when 1-3 items are presented, enumeration is fast, accurate and relatively independent of number – a process termed *subitizing*. Larger sets decrease enumeration accuracy and increase response times (Trick and Pylyshyn 1994; Railo et al. 2016). The number of items that can be "subitized" is considered to mark the capacity of items that can be individuated and attended in a single glance (Jevons 1871; Trick and Pylyshyn 1994; Railo et al. 2008; Piazza et al. 2011). If binocular disparity increases the speed and capacity of perception, enumeration should become faster and more accurate when compared to the condition where visual disparity has been removed (the same image is presented to both eyes). The present study employed very brief presentation times (50 msec) combined with backward masking to focus on the contribution of information contained in visual disparity per se. Such brief stimulus durations minimize the contributions from motion parallax (if the person moves his head), but also limits the strength of subjective perception of depth (Westheimer 1994). However, if visual disparity contributes to the

speed and accuracy of segmenting objects from the background, the effect should be observed even with short presentation times that may be insufficient to produce a strong subjective sensation of depth.

Method

Participants

Twenty-nine participants took part in Experiment 1 (age range 18-34, 23 females), and 31 participants to Experiment 2 (age range 19-35, 26 females). All participants reported having normal or corrected-to-normal vision. The study was conducted in accordance with the Declaration of Helsinki.

Stimuli and procedure

Participants were presented with stereoscopic and non-stereoscopic stimuli in immersive virtual reality using the Oculus Rift DK2 headset (resolution 960 × 1080 per eye, 75 Hz refresh rate, 100° field of view). Immersive virtual reality refers to the fact that the participant feels being present in the virtual world. The world is immersive in the sense that it surrounds the participant and by turning her head the participant can look around in the virtual world. The Oculus Rift was running on a computer (3.50 GHz, 16 GB RAM) with a NVIDIA GeForce GTX 980 graphics card. The camera that tracks the position and orientation of the headset was placed roughly on the table roughly 1 meter from the participant.

In Experiment 1 the stimuli were natural scenes that contained 1–6 humans (see Fig. 1). The humans were generated using an open source Make Human software (www.makehuman.org). They were distributed around the scene, they never overlapped with each other, and at least their head and torso were clearly visible. The scenes, which were generated using various “prefabs” of the Unity software, contained trees, bushes or rocks, but never any artificial, man-made objects. The humans were positioned directly in front of the camera. In Experiment 2 the stimuli were randomly oriented

cubes presented at pseudo-random positions in front of the participant on sky background (same as in Experiment 1). The color of the cubes was constant within a scene, but varied across scenes.

The experiment and the stimuli were made using the game engine Unity (5.4.0f3). The images presented to the left and right eye screens in the Oculus Rift were determined by two cameras placed in the virtual world constructed by Unity. We used the default settings: the distance between the left and right virtual cameras was 64 mm, and they were parallel to each other, that is, the cameras were not angled inward like eyes typically are. While this means that the size of visual disparity is reduced (which in turn is likely to reduce the size of the effect disparity has on enumeration performance), the advantage is that the scenes with and without visual disparity look very similar (the manipulation of disparity was not spontaneously noticed by the participants). This is important as strong variation in disparity could draw the participants' attention and influence behavior indirectly. Furthermore, stronger disparity could produce nausea in some participants. In the "No disparity" condition the image of the left virtual camera was presented to both left and right eyes of the participant.

As shown in Figure 1, each trial began with the presentation of a light gray square positioned directly in front of the participants. When the participant oriented towards the square for 1 second it turned to a darker shade and triggered the presentation of the scene 700 msec later. Each scene was presented for 50 msec, and it was followed by a flat (i.e. 2D) pixelated mask image for 200 msec. In the enumeration task the participant was asked to say out loud the number of humans presented as quickly as possible. The participants wore Sennheiser PC-8 headphones with a microphone. Reaction times relative the onset of the scene were recorded. After the vocal response numbers 1–9 were presented on a 3 x 3 grid of squares in the virtual reality. The participant selected the number she had spoken by orienting towards it in the virtual reality. The answer was selected when the participant kept orienting towards the number for at least 1 second. In the subjective experience of depth task the participant was asked to rate the strength of the subjective experience of depth and three-dimensionality on a scale from one to three. The highest alternative denoted a strong sense three-dimensionality, whereas the lowest alternative meant that the participant experienced a relatively flat, 2D world with very little sensation of depth. The middle alternative referred to medium strength experience of depth. We also emphasized to the participant that this task was not related to the apparent quality of the graphics, nor was there a correct answer to the question. The participant was

told that she did not have to make a speeded response. The answer was given first by speaking to a microphone and then selecting the corresponding response by orienting to it in the virtual reality.

Each participant performed two blocks of 60 trials of the enumeration and two blocks of 60 trials of the depth rating task. The order of the tasks was counterbalanced across participants. Each participant saw each stimulus once, and across participants each stimulus appeared equally often in the enumeration and in the depth rating tasks. Ten practice trials were completed before each task.

Statistical analysis

Data were analyzed in R statistical software (version 3.1.2; (R Development Core Team 2014)). Single-trial depth rating and reaction time data was analyzed using linear mixed-effects models (Baayen, Davidson, and Bates 2008), and accuracy data with logit mixed-effects models (Jaeger 2008) with binomial probability distributions (using the lme4 package; (Bates et al. 2014)). Mixed-effects models enable taking into account individual differences (e.g., in the speed of responding) in the model, in addition to group-level effects. In the mixed-effect models, the number of items (1–6), binocular disparity (disparity vs. no disparity), and their interaction were added as fixed-effects factors. In addition, the reaction time of a previous trial was included to reaction time models, and the depth rating of a previous trial was included to the depth rating analyzes. Both of these continuous variables were centered. Participants were defined as random effects so that each participant's data were fitted with individual intercepts and slopes (for both disparity and number conditions). Moreover, because enumeration data is known to contain a breakpoint around number 2-4, the data was fitted with two-segment linear models in addition to standard one segment models (for formal description, see (Railo et al. 2016)). We fitted three alternative two-segment models where the length of the first segment ranged numbers 1–2, 1–3, or 1–4. These were compared against the one segment model with a corresponding random-effect structure. Model fits were evaluated using the Akaike Information Criterion, which weighs the goodness of fit by the complexity of the model. Different models were compared using likelihood-ratio tests. The best fitting model was pruned by removing nonsignificant regressors ($t < |2|$). The pruned models are visualized in the figures. We do not report p values because there is no standardized way to calculate the degrees of

freedom and the corresponding p values in mixed-effects models (Baayen, Davidson, and Bates 2008). The statistical significance of an effect can be checked from the confidence intervals (CI): whenever the value zero is included in the CI, the effect is not considered statistically significant. Error bars in all the figures are 95% CIs calculated from 1000 bootstrap samples.

Datasets and the analysis scripts can be downloaded at the Open Science Framework (https://osf.io/9b3h8/).

Results

Experiment 1: Natural scenes

The results of Experiment 1 where the participants viewed natural scenes are presented in Figure 2. At group level, subjective depth ratings decreased .20 units on average when binocular disparity was removed (β = -0.20, CI= [-0.25, -0.15], t = 8.5). Moreover, depth ratings increased with number (β = 0.065, CI = [-0.043, 0.087], t = 5.9). The interaction between number and disparity condition was not statistically significant, and was thus left out of the model. Finally, the depth rating of previous trial predicted ratings of depth (β = -0.089, CI = [0.058, 0.12], t = 5.6).

Enumeration accuracy was best fitted with a two segment model (χ^2 = 12.1, df = 4, p = .016): Accuracy remained high for the first three items (subitizing), after which it declined rapidly. Accuracy was slightly lower without binocular disparity during the first segment (OR = 0.66, CI = [0.49, 0.90], Z = -2.61). Accuracy also slightly decreased as a function of number for the first three numbers (OR = 0.45, CI = [0.38, 0.55], Z = -8.04). In the second segment, enumeration accuracy declined more steeply as a function of number (OR = 0.54, CI = [0.40, 0.73], Z = -3.9). An interaction between disparity condition and number (4-6) in the second segment indicated that enumeration accuracy was initially higher in the second segment during binocular disparity (OR = 1.44, CI = [1.13, 1.86], Z = 3.00). This result is important as it shows that during binocular disparity enumeration remains accurate longer than without disparity. In addition to modelled results, this result is also clearly seen in the observed data: Bonferroni corrected pairwise t tests between the disparity conditions at each number

confirmed that a statistically significant difference was only present at number four (difference 8.5 percentage units, p = .002; other comparisons: p values ≥ .37).

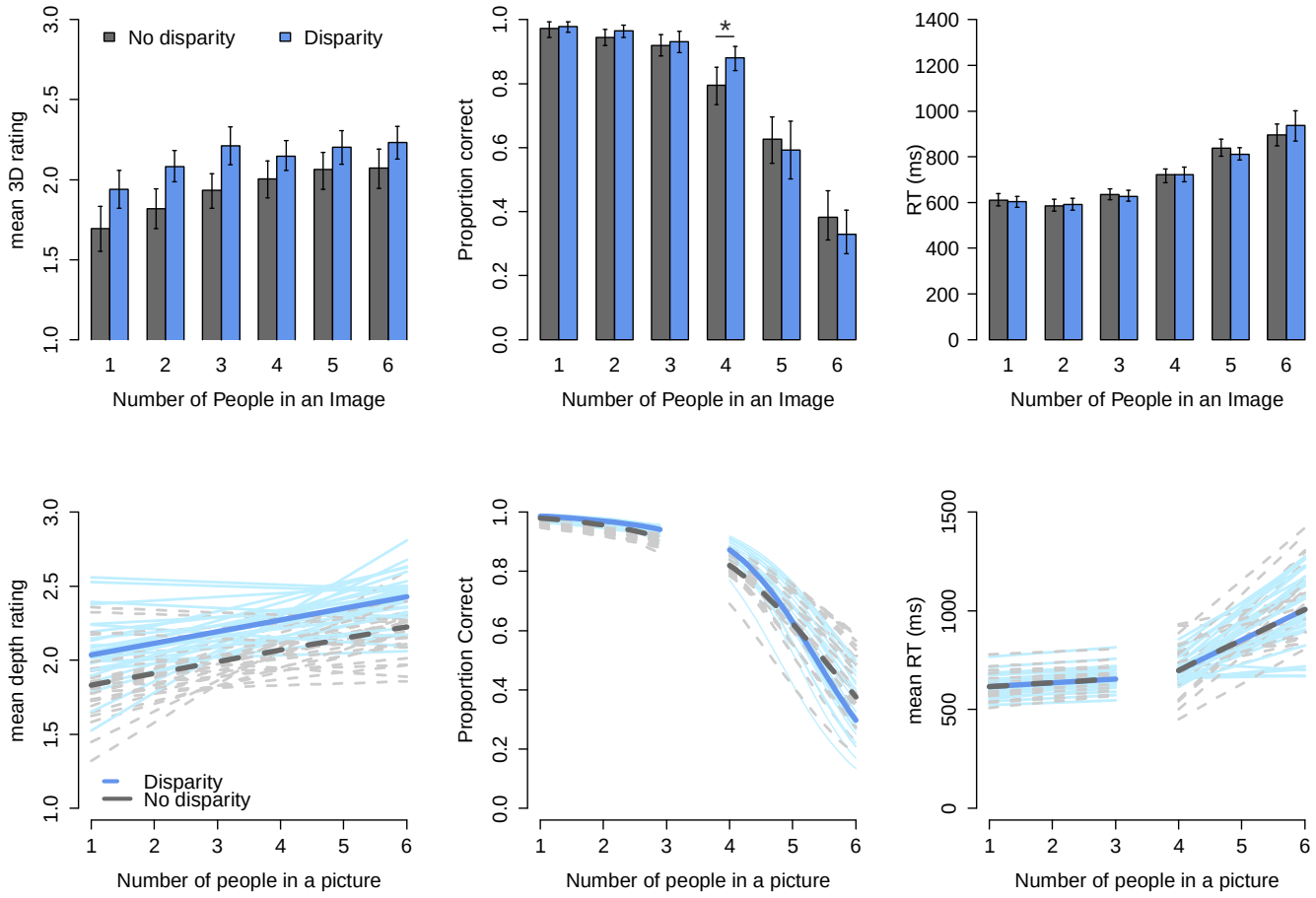

Figure 2. Results of Experiment 1: Observed data (upper panels) and modelled data (lower panels) when participants viewed natural scenes. Error bars show 95% CIs. In the lower panels, the bold lines depict the group-level results and the thinner lines in the background show individual participants' results (i.e. random-effect variation).

Reaction times were also best modelled by a two-segment model rather than a corresponding one segment linear model (χ^2 = 354.1, df = 37, p < .0001). Reaction times slightly increased for the first three numbers (β = 13.14 ms, CI = [6.46, 19.81], t = 3.86; intercept: β = 616.03 ms, CI = [596.01, 636.06], t = 60.30). For the second segment enumeration times increased more steeply (β = 93.46 ms,

CI = [73.299, 105.62], t = 10.85). Thus, whereas binocular disparity increased enumeration accuracy, it did not influence the speed of correct responses.

Experiment 2: Cube stimuli

Experiments 1 suggests that when viewing natural scenes binocular disparity may increase visual capacity without influencing the speed of visual processing. Experiment 2 was designed to test whether these effects are larger when the stimuli contain very few other depth cues than binocular disparity. In Experiment 2 we employed geometric cubes presented against a distant sky background as the stimuli to be enumerated (see Figure 1). Here other depth cues (than binocular disparity) are minimized whereas the natural scene stimuli employed in Experiment 1 contains various depth cues (e.g. familiar objects in predictable environments, occlusion, shading). We expected this manipulation to enhance the effect binocular disparity has on enumeration performance.

The results are shown in Figure 3. Depth ratings were best fitted by a two-segment model where the first segment included numbers 1-3 (χ^2 = 27.3, df = 4, p < .0001). Surprisingly, participants' subjective experience of depth was not modulated by binocular disparity. For numbers 1-3, mean depth rating was 1.90 (β = 1.90, CI = [1.80,2.00], t = -37.6), and for numbers 4-6 depth ratings increased by 0.24 (β = 0.24, CI = [0.18, 0.31], t = -7.3).

The accuracy of enumeration in Experiment 2 was very similar to the results obtained in Experiment 1 with day time scenes as stimuli. Enumeration accuracy was best fit by a two-segment model (χ^2 = 15.1, df = 1, p < .0001). In the first segment, enumeration accuracy decreased a function of number in the first segment (OR = 0.24, CI = [0.18, 0.39], Z = -9.6), and enumeration accuracy was lower when binocular disparity was eliminated (OR = 0.61, CI = [0.43, 0.87], Z = -2.7). Crucially, the second segment also revealed an interaction between number and disparity condition (OR = 1.42, CI = [1.09, 1.84], Z = -2.6): Enumeration remains accurate longer with binocular disparity. Bonferroni corrected t tests at each number comparing the two disparity conditions verified that enumeration accuracy was reduced without binocular disparity (difference 14.3 percentage units; p = .0048) only at

number four (other uncorrected p values ≥ .48). All in all, these results with cube stimuli replicate the results obtained with day time scenes concerning enumeration accuracy.

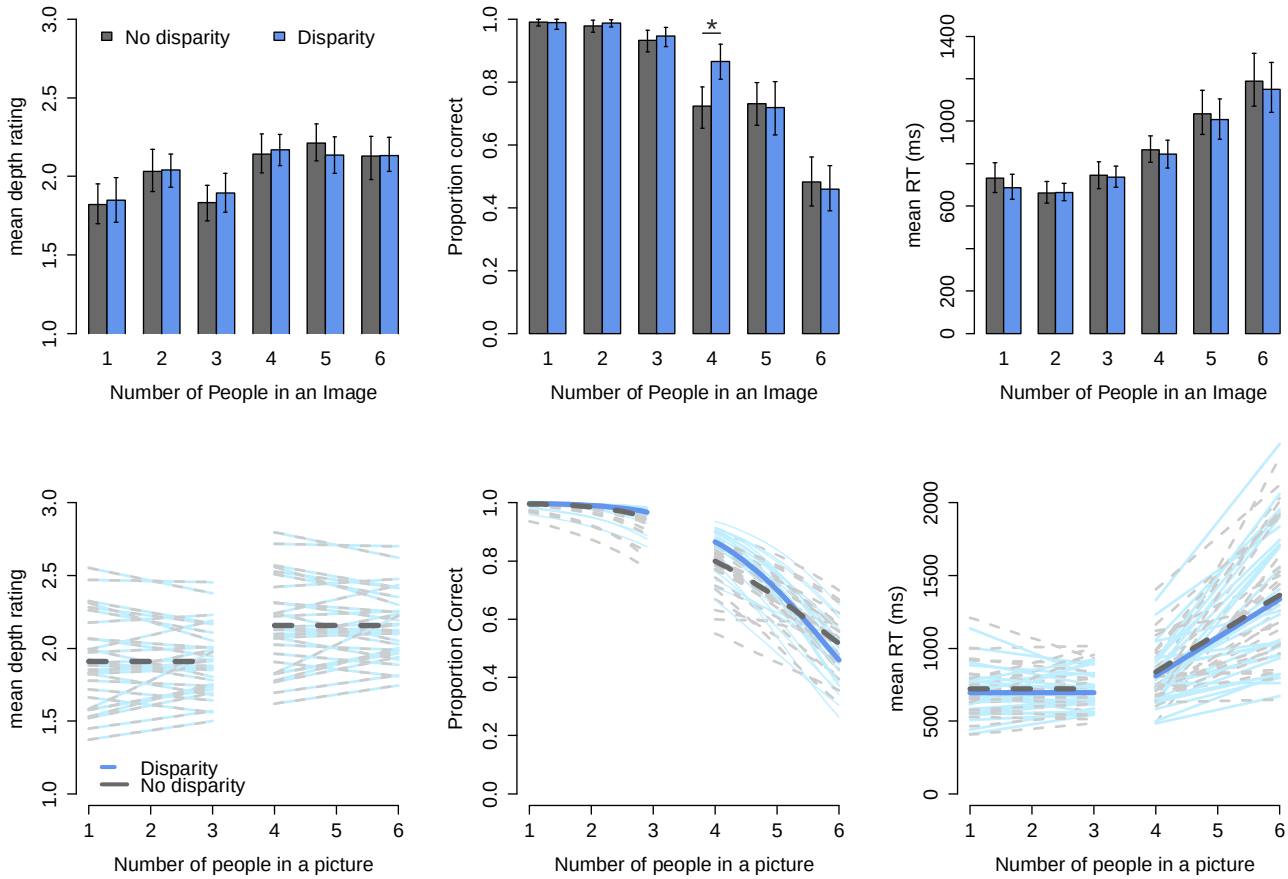

Figure 3. Results of Experiment 2: Observed data (upper panels) and modelled data (lower panels) when participants viewed geometric shapes. Error bars show 95% CIs. In the lower panels, the bold lines depict the group-level results and the thinner lines in the background show individual participants' results.

In Experiment 2 reaction times were delayed by a constant of 27 ms when binocular disparity was removed (see Fig 3; β = 27.04 ms, CI = [8.08, 45.99], t = 2.7). That is, with cubes as stimuli, enumeration was *faster and more accurate* with disparity than without it. Otherwise, the reaction times followed a similar two-segment pattern (χ^2 = 310.2, p < .0001) as previously. In the first segment reaction times remained constant (β = 693.2 ms, CI = [647.66, 738.81], t = 29.8). During the second segment reaction times increased as a function of number (β = 175.8 ms, CI = [130.71, 220.82], t = 7.6).

Discussion

Different theories of stereopsis have in common the assumption that frontal eyes and the resulting binocular disparity represents an adaptation that enables the animals to better discriminate objects from the background (Crompton 1995). The present results show that even very slight binocular disparity, one that is not noticed by the participant and does necessarily produce a concurrent enhancement of consciously perceived depth, increases the capacity and speed to enumerate objects in realistic natural scenes. This supports the idea that during primate evolution, stereopsis helped to provide the animals with more accurate and richer visual percepts. This may have helped the animals to better detect objects (e.g. fruit, prey, predators) that are embedded among other objects (e.g. plants and branches) in a complex visual environments (Sussman 2013). Similarly, faster visually guided behavior may have helped the animals to catch prey (Julesz 1971; Cartmill 1974) and escape predators (Isbell 2009). During evolution, the possibility of better visual discrimination and faster visual behavior provided by binocular disparity may have been an important factor driving the increase in brain size, especially the expansion of visual cortical areas in primates (Barton 2004). Our results also raise the possibility that individuals with impaired binocular vision have a reduced visual perceptual capacity.

With binocular disparity, the accuracy of enumeration was increased most notably when the scene contained four target individuals, indicating that binocular disparity enhanced the accuracy of perception right at the limit of the subitizing capacity. Disparity increased enumeration accuracy on average by 8.5 and 14.2 percentage units in Experiments 1 and 2, respectively. The size of our results are similar to the improvement observed when experienced video-game players are compared to non-gamers – as in the present study, the observed increase in enumeration accuracy was about 10 percentage units, and began at number four (Green and Bavelier 2003; Green and Bavelier 2006). Note that although the modelled results suggest that binocular disparity may decrease enumeration accuracy at high numbers this effect did not reach statistical significance in the observed data. Hence, this effect may represent a byproduct of the (nonlinear) logit model used to model the accuracy data.

When geometric cube stimuli were used, disparity also sped-up reaction times by 27 ms on average. While the absolute size of this effect on reaction times is relatively small, it may represent a

significant advantage in situations where fast visually-guided behavior is important. The reader should also note that logically one could have also expected to observe longer reaction times under binocular disparity, as integrating the two different viewpoints of the eyes is an complicated task that can be assumed to require additional processing time. Reaction times were enhanced only in Experiment 2 possibly because other depth cues were minimized with the use of geometric shapes.

There is reason to assume that the size of the advantage produced by binocular disparity would be larger if the target stimuli were closer to the observer (producing larger disparity), if binocular disparity would be increased by aligning the virtual cameras (i.e. eyes) inward (producing larger disparity), and if the scene would be presented for a longer duration (enabling longer accumulation of visual information). We purposely limited the amount of disparity and presentation duration so that the participants would not notice the experimental manipulation (change in disparity). Strong variation in disparity could make the participant notice the manipulation, and draw their attention to it, and thereby influence behavior indirectly. None of the participants in the present experiments reported noticing the disparity manipulation.

We propose that superior accuracy in enumerating four objects with disparity is produced by the fact that disparity enabled the participants to better individuate the objects from each other objects and the background. Franconeri et al. (2013) have suggested that capacity limits in perception result from competition between objects in representational maps. Binocular disparity may counteract the competition between different objects within representational maps, enabling the visual system to more efficiently process the objects. This advantage is small when visual discrimination is relatively easy (i.e. within the subitizing range), but the advantage produced by disparity becomes apparent in situations where the number of items is just above the limits of visual processing capacity (i.e. four items). This effect could arise at earliest in V1, which is the first stage of visual processing that contains disparity selective neurons, although higher cortical areas may also contribute to the effect (Barlow, Blakemore, and Pettigrew 1967; Cumming and DeAngelis 2001). Stereoscopic perception constrains the allocation of selective visual attention (He and Nakayama 1995), and as subitizing is known to depend on attentional resources (Railo et al. 2008), the better enumeration performance under binocular disparity could be due to easier deployment of attention to the objects to be enumerated.

Interestingly, our results show that the advantage produced by binocular disparity does not require an enhanced conscious perception of depth, suggesting that the mechanisms that produce the functional advantage are somewhat independent from the processes that produce the conscious experience of depth often associated with disparity. The dissociation between enhanced subjective perception of depth and the functional advantage provided by binocular disparity is in concert with recent theoretical proposals that emphasize the difference between functional and subjective forms of stereopsis (Vishwanath 2014), and with the fact that binocular disparity is not necessary for subjective sensation of depth (Vishwanath and Hibbard 2013). Moreover, binocular disparities can elicit vergence eye movements in the absence of subjective perception of depth (Masson, Busettini, and Miles 1997). While vergence eye movements and the behavioral effects observed in the present study could be mediated by disparity-selective neurons in V1, these neurons are unlikely to account for subjective perception of depth (Cumming and Parker 1997). We suggest that while the functional advantage produced by binocular disparity is likely produced by following rapid computations (possibly in V1; (Masson, Busettini, and Miles 1997; Cumming and Parker 1997)), the encoding of subjective experience of depth requires longer processing times and integration of information across different cortical areas (Koch et al. 2016). Further studies are required to pinpoint the neural mechanisms involved in the two processes. For instance, whereas the functional advantage provided by disparity could also be mediated predominantly by fast acting dorsal visual stream/magnocellular neurons, but the ventral stream/parvocellular neurons may contribute more to the subjective experience of depth (Parker 2007).

While one of the aims of the present study was to examine whether binocular disparity aids perception, we do not wish to imply that perception and behavior are two completely separate processes. Although the present experiment did not require visuomotor behavior in the sense that the participant did not need to reach for any objects, the fast spoken response (employed in the present study) is nevertheless a form of visually-guided behavior. It is likely that the task performed by the participants influences the advantage provided by binocular disparity. For instance, tasks that require the observer to infer or utilize spatial relations of objects may be more sensitive to the advantage provided by binocular disparity. It is likely that the spatial layout of the objects may influence the advantage provided by binocular disparity. Cluttered scenes where the objects are close

to the observer (i.e. resulting in large disparity) are likely produce a larger advantage. Also, it remains to be tested whether the visual system is more sensitive to binocular depth cues when, for example, either threatening or otherwise "evolutionarily relevant" (i.e. snakes, see (Isbell 2009)) are present.

In conclusion, the present results suggest that binocular disparity enhances the capacity of visual behavior by increasing the accuracy and speed at which multiple objects can be individuated from the background. This effect can arise without concurrent enhancement in consciously experienced depth. During evolution, the computations required to resolve binocular disparity may have been important in driving the expansion of cortical areas devoted to visual processing to extract a richer visual perception. The higher capacity of visual perception may have enabled more efficient visual behavior, as optimal performance in complex visual environments requires accurate segregation of not just one but few objects simultaneously.

Acknowledgements

The research was supported by the Finnish Cultural Foundation. We thank Iiris Ollikainen for help with data collection.

Competing interests

No competing interests declared

References

Baayen, R H., D J. Davidson, and D M. Bates. 2008. "Mixed-effects modeling with crossed random effects for subjects and items." *Journal of Memory and Language* 59:390-412.

Barlow, H B., C Blakemore, and J D. Pettigrew. 1967. "The neural mechanism of binocular depth discrimination." *The Journal of Physiology* 193(2):327-342.

Barry, S R. 2009. *Fixing My Gaze*. New York, NY: Basic Books.

Barton, R A. 2004. “Binocularity and brain evolution in primates.” *Proceedings of the National Academy of Sciences of the USA* 101(27):10113-10115.

Bates, D, M Maechler, B M. Bolker, and S Walker. 2014. “lme4: Linear mixed-effects models using Eigen and S4.” *R package version 1.1-7, URL: http://CRAN.R-project.org/package=lme4*.

Cartmill, M. 1974. “Rethinking primate origins.” *Science* 184:436-443.

Cartmill, M. 1992. “New views on primate origins.” *Evolutionary Anthropology* 1:105-111.

Collins, E T. 1921. “Changes in the visual organs correlated with the adoption of arboreal life and with the assumption of the erect posture.” *Transactions of the Opthalamological Society of the UK* 41:10-90.

Crompton, R H. 1995. “"Visual predation," habitat structure, and the ancestral primate niche.” Pp. 11-30 in *Creatures of the Dark: The Nocturnal Prosimians*, edited by L Alterman, G A Doyle, and M K Izard. New York, New York: Plenum Press.

Cumming, B G. and G C. DeAngelis. 2001. “The physiology of stereopsis.” *Annual Review of Neuroscience* 24:203-238.

Cumming, B G. and A J. Parker. 1997. “Responses of primary visual cortical neurons to binocular disparity without depth perception.” *Nature* 389:280-283.

Dal Mutto, C, P Zanuttigh, G M. Cortelazzo, and S Mattoccia. 2011. “Scene segmentation assisted by stereo vision.” *International Conference on 3D Imaging, Modeling, Processing, Visualization, and Transaction* 57-64.

Fielder, A R. and M J. Moseley. 1996. “Does stereopsis matter in humans.” *Eye* 10:233-238.

Franconeri, S L., G A. Alvarez, and P Cavanagh. 2013. “Flexible cognitive resources: competitive content maps for attention and memory.” *Trends in Cognitive Sciences* 17(3):134-141.

Green, C S. and D Bavelier. 2003. “Action video game modifies visual selective attention.” *Nature* 423:534-537.

Green, C S. and D Bavelier. 2006. “Enumeration versus multiple object tracking: The case of action video game players.” *Cognition* 101:217-245.

Heesy, C P. 2009. “Seeing in stereo: The ecology and evolution of primate binocular vision and stereopsis.” *Evolutionary Anthropology* 18:21-35.

He, Z and K Nakayama. 1995. “Visual attention to surfaces in three-dimensional space.” *Proceedings of the National Academy of Sciences of USA* 92:11155-11159.

Isbell, Lynne. 2009. *The Fruit, the Tree and the Serpent*. Cambridge, Massachusetts: Harvard University Press.

Jaeger, T F. 2008. “Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models.” *Journal of Memory and Language* 59(4):434-446.

Jevons, W S. 1871. “The power of numerical discrimination.” *Nature* 3:281-282.

Jones, F. 1917. *Arboreal Man*. New York, NY: Hafner Publishing.

Julesz, B. 1971. *Foundations of Cyclopean Perception*. Cambridge, Massachusetts, London, England: MIT Press.

Koch, C, M Massimini, M Boly, and G Tononi. 2016. "Neural correlates of consciousness: progress and problems." *Nature Reviews Neuroscience* 17(5):307-321.

Masson, G S., C Busettini, and F A. Miles. 1997. "Vergence eye movements in response to binocular disparity without depth perception." *Nature* 389:283-286.

McKee, S P., N J. Watamaniuk, J M. Harris, H S. Smallman, and D G. Taylor. 1997. "Is stereopsis effective in breaking camouflage for moving targets." *Vision Research* 37(15):2049-2055.

Melmoth, D R., A L. Finlay, M J. Morgan, and S Grant. 2009. "Grasping deficits and adaptations in afults with stereo vision losses." *Investigative Ophalamology & Visual Science* 50(8):3711-3720.

O'Connor, A R., E E. Birch, S Anderson, H Draper, and FSOS Research Group. 2010. "The functional significance of stereopsis." *Investigative Opthalamology & Visual Science* 51(4):2019-2023.

Parker, A J. 2007. "Binocular depth perception and the cerebral cortex." *Nature Reviews* 8:379-391.

Pettigrew, J D. 1986. "The evolution of binocular vision." Pp. 208-222 in *Visual Neuroscience*, edited by J D Pettigrew, K J Sanderson, and W R Levick. London: Cambridge University Press.

Piazza, M, A Fumarola, A Chinello, and D Melcher. 2011. "Subitizing reflects visuo-spatial object individuation capacity." *Cognition* 121:147-153.

R Development Core Team. 2014. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org/.

Railo, H, V.-M. Karhu, J Mast, H Pesonen, and M Koivisto. 2016. "Rapid and accurate processing of multiple objects in briefly presented real-world images." *Journal of Vision* 16:doi:10.1167/16.3.8.

Railo, H, M Koivisto, A Revonsuo, and M M. Hannula. 2008. "The role of attention in subitizing." *Cognition* 107:82-104.

Ross, C F. and R D. Martin 2006. "The role of vision in the origin and evolution of primates." Pp. 59-78 in *Evolution of Neurvous Systems*, edited by J H Kaas. Oxford: Academic Press.

Saint-Amour, P. 2003. "Modernist Reconnaissance." *Modernism/modernity* 10(2):349-380.

Smith, G E. 1924. *The Evolution of Man*. London: Oxford University Press.

Sussman, R W. 1991. "Primate origins and the evolution of angiosperms." *American Journal of Primatology* 23:209-223.

Sussman, R W. 2013. "Rethinking primate origins again." *American Journal of Primatology* 75:95-106.

Trick, L M. and Z W. Pylyshyn. 1994. “Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision.” *Psychological Review* 101(1):80-102.

Wardle, S G., J Cass, K R. Brooks, and D Alais. 2010. “Breaking camouflage: Binocular disparity reduces contrast masking in natural images.” *Journal of Vision* 10(14):1-12.

Westheimer, G. 1994. “The Ferrier Lecture, 1992. Seeing depth with two eyes: stereopsis.” *Proceedings of the Royal Society B: Biological Sciences* 257:205-214.

Vishwanath, D. 2014. “Toward a new theory of stereopsis.” *Psychological Review* 121:151-178.

Vishwanath, D and P B. Hibbard. 2013. “Seeing in 3-d with just one eye: stereopsis without binocular vision.” *Psychological Science* 24(9):1673-1685.