As accepted for publication in Developmental Science

Does culture shape face perception in autism? Cross-cultural evidence of the own-race advantage from the UK and Japan

Mary Hanley[1, 6], Deborah M Riby[1, 6], Michael-John Derges[1], Anna Douligeri[1], Zackary Philyaw[1], Takahiro Ikeda[2], Yukifumi Monden[2,3], Hideo Shimoizumi[4], Takanori Yamagata[2], Masahiro Hirai[5, 6].

[1]Department of Psychology, Durham University, UK

[2] Department of Pediatrics, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan

[3] International University of Health and Welfare Hospital Address: 537-3 Iguchi, Nasushiobara, Tochigi 329-2763 Japan

[4] International University of Health and Welfare Rehabilitation Center, Nasu Institute for Developmental Disabilities, 2600-7 Kitakanamaru, Otawara, Tochigi, 324-0011, Japan

[5] Center for Development of Advanced Medical Technology, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke, Tochigi 329-0498, Japan

[6] Centre for Developmental Disorders, Durham University, UK

Correspondence to: Professor Deborah M Riby, Department of Psychology, Durham University, Science Site, South Road, Durham DH1 3LE. Telephone: 01913343247. Email: deborah.riby@durham.ac.uk

Conflict of Interest: The authors report no conflicts of interest.

Acknowledgements: The authors wish to acknowledge the support of all schools and families that participated in this research. This collaboration was supported by the British Psychological Society Developmental Section's International Collaboration Scheme award to Dr Mary Hanley. This work was also supported by a JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 18H01103, Grant Number 23830127, the grant in aid for Scientific Research on Innovative Areas 15H01585 'Constructive Developmental Science' from the Ministry of Education, Culture, Sports, Science and Technology, and Pfizer health research foundation, which was awarded to Masahiro Hirai. We would like to thank Dr Holger Wiese for his helpful discussions concerning data analysis and interpretation, and Ali Williams for her contribution to data collection for part of the sample. Finally, we wish to thank Sarina Hui Lin Chien and her co-authors (Chien et al., 2014) for sharing the experimental stimuli.

Research Highlights

- The first study to investigate the own-race advantage in Japan and the UK simultaneously
- The presence of own-race advantage is heterogeneous in both autism and typical development
- Children with autism in the UK and Japan showed typical face recognition, even when requiring manipulations to the eye regions
- Atypical experience with faces in autism does not lead to a reduced/absent own-race advantage

Abstract

Autism spectrum disorders (ASD) are associated with face perception atypicalities, and atypical experience with faces has been proposed as an underlying explanation. Studying the own-race advantage (ORA) for face recognition can reveal the effect of experience on face perception in ASD, although the small number of studies in the area present mixed findings. The current study probed the ORA in ASD by comparing two cultural groups simultaneously for the first time. Children with ASD in the UK (N=16) and Japan (N=26) were compared to age and ability matched TD children in the UK (N=16) and Japan (N=26). Participants completed a two-alternative forced-choice task, whereby they had to recognise a just-seen face from a foil which was manipulated in one of four ways (IC: identity change; EE: easy eyes; HE: hard eyes; HM: hard mouth). Face stimuli were Asian and Caucasian, and thus the same stimuli were own and other-race depending on the cultural group. The ASD groups in the UK and Japan did not show impaired face recognition abilities, or impairments with recognising faces depending on manipulations to the eye region, and importantly they showed an ORA. There was considerable heterogeneity in the presence of the ORA in ASD and TD and also across cultures. Children in Japan had higher accuracy than children in the UK, and TD children in Japan did not show an ORA. The present cross-cultural study challenges the view that atypical experiences with faces lead to a reduced/absent ORA in ASD.

Key words

Autism Spectrum Disorders, own-race effect, cross-cultural, face recognition

Does culture shape face perception in autism? Cross-cultural evidence of the own-race effect from the UK and Japan

A wealth of evidence supports the idea that faces represent a special class of visual stimuli (Park, Newman & Polk, 2009). Faces capture our attention (Langton, Law, Burton, & Schweinberger, 2008), we spend longer looking at them than other types of visual stimuli (Birmingham, Bischof & Kingstone, 2008), and we develop dedicated neural networks for processing facial cues (Birmingham & Kingstone, 2009; Park et al., 2009). Indeed it has been proposed that the Fusiform Face Area (FFA) of the brain is dedicated to processing this special class of stimuli (e.g. Kanwisher, McDermott, & Chun, 1997). Faces are especially important because they convey social information that guides inter-personal communication, for example regarding a person's identity (face recognition), how they might be feeling (emotion recognition), and what they might be thinking (mental state attribution). Therefore, it is not surprising that faces hold a special status for our attention, given that they provide us with crucial information for social perception and cognition.

This is evident even in early infancy, when despite relatively poor visual acuity, infants have shown sensitivity to face stimuli (Johnson, Dziurawiec, Ellis, & Morton, 1991). For example, they prefer to attend to face-like stimuli over non-face stimuli, and they prefer familiar to unfamiliar faces (Pascali et al., 2011). However, face perception ability is far from mature in infancy and follows a protracted period of development (Bruce et al., 2000; Chein, Tai, & Yang, 2018; Mondloch, Le Grand & Maurer, 2002; but see McKone, Kanqisher & Duchaine, 2007). We know that our sensitivity to faces is shaped in important ways by experience, for example by the people we spend most time with (usually members of our 'own-race') and by our culture. For example, avoidance of eye contact is considered a sign of respect in Eastern cultures, but not within Western cultures (Sue & Sue, 1977). Related to this, participants from Eastern cultures have reported direct eye contact can lead to judgements of increased anger, reduced approachability and increased unpleasantness, compared to Western participants (Akechi et al., 2013). Furthermore, evidence has indicated that basic perceptual processing of faces may be subject to cultural influence, with a holistic processing style (i.e. context focussed) characterising perceptual skills in Eastern cultures compared to an analytic style (i.e. focal information focussed) in Western cultures (Blais, Jack, Sheepers, Fiset & Caldara, 2008). Beyond attitudes and cultural influence, our experience with members of our cultural group actually influences our face recognition ability. Studies of the own-race advantage (ORA) show that even by mid-to-late childhood we are typically better at recognising unfamiliar faces from our own race than unfamiliar faces from other races (Anzures et al., 2014; Chien, Tai & Yang., 2018; Meissner & Brigham, 2001). Comparing different cultures offers the unique opportunity to probe the development of face perception abilities and the relative contribution of experience. This is particularly important when trying to understand groups for whom face perception might develop in an atypical manner, and the focus of the current study is Autism Spectrum Disorder (ASD); a group for whom this is the case.

The Development of the ORA in Children

Existing research has demonstrated the ORA (Meissner & Brigham, 2001) both in adults (Caucasian and Asian; O'Toole, Deffenbacher, Valentin & Abdi, 1994) and in children (Caucasian and African-American; Pezdek, Blandon-Gitlin, & Moore, 2003) showing that a bias for recognising own-race faces is present cross-culturally and relatively early in development. Research has shown the developmental origins of the ORA exist during infancy, but that experience is crucial. At three months old, Caucasian infants were able to discriminate between different exemplars of Caucasian, African and Chinese faces using a novelty preference paradigm, whilst at six months they could only do so for Caucasian and Chinese faces, and at nine months only Caucasian faces (Kelly et al., 2007a). This emphasises the important experience dependent developmental changes that occur in relation to the ORA, even by 9 months of age. [1]

Beyond infancy, the size of the ORA gradually increases throughout childhood (Chance, Lockwood Turner & Goldstein, 1982; Chien, et al., 2018; Goodman, Hirschman, Hepps & Rudy, 1991; Goodman et al., 2007). Although some have argued that it is stable by 5 years of age (Anzures et al., 2014; de Heering, de Liedekerke, Deboni, & Rossion, 2010; Sangrigoli, Pallier, Argenti, Ventureyra, & de Schonen 2005), the balance of evidence indicates that an adult-like ORA is present towards the latter end of childhood. For example, Chance et al. (1982) studied the development of the ORA from childhood to adulthood with participants ranging in age from 6-20 years. They used an old/new face recognition task with both Caucasian and Japanese faces, and found that although accuracy generally improved for both types of faces with age, the youngest children (7-8 years) did not show an ORA, whereas the older children (11-12 years) and adults did. Chien et al. (2018) used a face morphing paradigm with a sample of 100 5-12 year old children and 23 adults from Taiwan. The task involved same/different judgements on faces which were either own-race (Asian – parent condition) or other race (Caucasian – parent condition) in origin, and which were

[1] While the Kelly et al. (2007a) study supports the perceptual narrowing view of the emergence of the ORA, there is debate in the literature as to whether perceptual narrowing or perceptual learning best explains the development of the ORA in infancy (for perceptual learning accounts see: Chien, Wang, & Huang, 2016; Hayden, Bhatt, Joseph, & Tanaka, 2007; Sangrigoli & de Schonen, 2004). This debate is beyond the scope of this study, particularly as the focus here is not on infancy, but simply the fact that experience with faces is key to shaping expertise (which both accounts support).

morphed to different degrees with the ‘other’ category (Asian faces morphed with Caucasian and vice versa). Across a number of measures (e.g. d’, rejection rates, threshold estimation) they found an absence of an ORA in 5-10 year olds, whereas the 11-12 year olds and adults showed an ORA in their better discrimination and sensitivity for the Asian-parent condition.

In sum, it is clear that the developmental origins of the ORA exist within infancy when experience with, and attention to, faces is important in shaping face perception. The ORA continues to develop over childhood, influenced by general improvements with face processing ability during this period linked with increasing experience with own-race faces. Therefore, for TD individuals, face expertise is biased towards those faces that are encountered most.

Face Expertise in Autism Spectrum Disorder

Although most typically developing individuals are considered to develop high levels of face expertise, the same is not true for individuals functioning on the autism spectrum. Autism spectrum disorders (ASDs) refer to a group of neurodevelopmental disorders characterised by deficits in social communication and the presence of restricted or repetitive patterns of behaviour, interests or activities (DSM-V, American Psychiatric Association, 2013). Difficulties with social communication are a defining feature of ASD and socio-communicative difficulties can be wide ranging (Rapin & Tuchman, 2008). Faces play a central role in social communication as emphasised above, and both difficulties processing information from others’ faces, and atypical face perception strategies, are prevalent in ASD (Dawson, Webb & McPartland, 2005; Nuske, Vivanti & Dissanayake, 2013; Sasson, 2006). It is worth noting that many of the face perception difficulties that have been reported in ASD (e.g. face recognition, face memory, emotion/mental state recognition; Fridenson-Hayo et al., 2016; Weigelt, Koldewyn & Kanwisher, 2013) have been associated with reduced and atypical attention to faces throughout development, especially reduced attention to the eye region (Corden, Chilvers & Skuse, 2008; Dalton et al., 2005; Spezio Adolphs, Hurley, & Piven, 2007). Adults, children and young toddlers with ASD have shown reduced and atypical attention allocation to faces, especially when there is competition from non-social information (Chita-Tegmark, 2016; Hanley et al., 2014; Jones et al., 2008; Klin, Jones, Schultz, Volkmar & Cohen, 2002). It has been suggested that reduced experience with faces from an early age may shape atypical face expertise and have a cascading impact upon the development of 'typical' face processing skills (Chawarska, Klin, Paul & Volkmar, 2007; Dawson et al., 2004; Chevalier et al., 2012).

One opportunity to understand the effect of reduced experience on face perception in ASD is to examine the ORA. If experience with faces is crucial to the development of the ORA as highlighted from the typical face perception literature, and reduced experience is important for shaping face perception in ASD, then we would expect children with ASD not to show a 'typical' advantage for recognition of own- compared to other-race faces.

The ORA and Autism

To date, five studies have investigated the ORA in ASD using face memory or face discrimination tasks within different cultural groups (using paradigms from the typical ORA literature). However, there are several inconsistencies in the methods and findings across these studies that have implications for our understanding of the presence and potential strength of an ORA in ASD.

Wilson et al. (2011) compared 27 children with ASD to 47 typically developing (TD) children (age range 6 to 16 years). This study of mostly Caucasian children (some were East Asian but had lived in Australia since birth) involved sequential two-alterntive forced-choice tasks with own (Caucasian) and other-race (Egyptian) face conditions. Overall, accuracy was better for the own compared to other-race faces, for both the ASD and TD groups suggesting an ORA in both typical and atypical development. However, there was vast heterogeneity in the ASD group and while those with age-appropriate face matching ability showed a typical ORA, those with lower face matching abilities did not show a typical ORA. So for those with ASD who have impaired face matching, the typical advantage for recognition of own-race faces does not develop and the authors proposed this is due to atypical social experience early in development which derails perceptual narrowing (Wilson et al., 2011).

Yi and colleagues (2015; 2016) conducted two studies on the ORA in ASD with Chinese participants and found inconsistent results. Yi et al. (2015) carried out a study with three groups of young adult Chinese participants (mean age 21 years), specifically ASD, TD, and those with intellectual difficulties (ID) who were matched to the ASD group based on non-verbal ability. Differing from Wilson et al. (2011), they used greyscale faces cropped to have all external features removed. On a face memory task the ASD and ID groups scored significantly better on the own-race faces compared to other-race faces, however the TD group showed comparable own and other-race face memory (and therefore no ORA). This could be taken as tentative evidence for an ORA in ASD. However, vastly different accuracy levels between the TD and ASD groups raised the question of whether differences in the task difficulty drove different performance patterns across the groups. In a follow-up study, Yi et al. (2016) compared Chinese children with ASD (age range 5-10 years) to age-matched TD and ability-matched TD children (age range 4.4 - 8.9 years), using the same method and stimuli as their previous study. They reported no difference in performance for the ASD children on own and other-race faces (e.g. no ORA). Both TD groups also showed comparable recognition of own and other-race faces (again no ORA). Considering Yi et al. (2015) and Yi et al. (2016) together, it is difficult to draw firm conclusions regarding the ORA in ASD, and it seems that the demands of the task are important for the interpretation of their results.

Recently, Hadad, Schwatrz and Binur (2019) provided stronge evidence of a reduced ORA in adults with ASD. Their study of adults with and without autism (mean age in years: TD 25.6 ; ASD 23.7) involved a face discrimination task with morphed face stimuli. In each trial, participants had to make a same/different judgement between a target and a morphed face (with either 20%, 40%, 60%, 80% or 100% of another face) presented simultaneously. Faces were either own-race (Caucasian) or other-race (Asian), and there was a further condition where same/different judgements were made on the same faces, but inverted. Their results indicated reduced specialisation for faces linked to experience, as they found a reduced ORA in adults with ASD driven by poorer sensitivity for recognition of own-race faces as well as reduced inversion effects for own-races faces.

The final study on the ORA in ASD has important methodological approaches that will be mirrored in the current study. Chien, Wang, Chen, Chen and Chen (2014) compared thirteen Taiwanese children with ASD to thirteen TD children matched for age (but not matched for verbal ability, ASD group had lower language ability; age range 6-10 years) and the task involved a two-alternative forced-choice paradigm (Chien et al., 2014). Children were presented with a target face, followed by the target and a foil face, and had to identify which face matched the one they had previously seen. The foil faces were manipulated in one of four ways to vary task difficulty: identity change (IC; different face); easy eyes (EE; same target face but with different eyes); hard eyes (HE; same target face but with eye spacing increased); and hard mouth (HM; same target face but with spacing between the mouth and nose decreased). Faces were either Asian or African. In terms of accuracy, there were differences between the groups for each condition, in that the TD group were better at the IC, EE and HE conditions, whilst the ASD group were better at the HM condition. This suggests that the ASD group struggled more with the easiest condition when discriminating between two completely different faces, and the two tasks where the critical information was located in the eye region.

Although no overall effects of face type or interactions with group were found, Chien and colleagues reasoned that the ORA may be subtle (alongside relatively small sample sizes for their main analysis) and so carried out independent sample t-tests separately for each of their task conditions. This revealed the TD group scored 100% on the own-race IC condition, whilst the ASD group scored significantly lower at 88%. In contrast when they compared performance for the TD and ASD groups on the other-race face IC condition, the groups performed similarly (the TD group scored 94%, whilst the ASD group scored 96%). They suggested that the difference on the own-race IC condition was indicative of a lack of experience with native faces. They explored this further by examining differential percent scores (own-race score minus other-race score). Positive scores indicated an own-race advantage, and a negative score an other-race advantage. On the easiest condition, the TD group had a significant positive score, whilst the ASD group a negative score (although not significantly different from zero). The authors suggested that on the IC condition the TD group demonstrated a significant ORA, whilst the ASD group did not show any difference.

In summary, the evidence regarding the presence or absence of an ORA in ASD is mixed. Chien et al. (2014) report that children with ASD lack the typical ORA, as do Hadad et al. (2019) for adults with ASD; Wilson et al. (2011) find this only in a sub-group of children with ASD with age inappropriate face recognition ability; Yi et al. (2016) report a lack of an ORA in ASD also, but at the same time, do not find it in TD children, while their study using the same task with adults with ASD shows a typical ORA at the same time that TD comparison adults do not. Some of these inconsistences may be due to the fact each study involved different age ranges (Wilson et al., 2001 6-16 years; Chien et al., 2014 6-10 years; Yi et al., 2016 5-10 years; Yi et al., 2015 mean 21 years; Hadad et al., 19-35 years). However, it is also pertinent to consider the methodological differences between studies. Some use two alternative forced-choice paradigms while others test face recognition memory. Of the two studies that use face recognition memory tasks, either the control group was at ceiling (Yi et al., 2015) or the ASD group was at chance level (Yi et al., 2016). Equally, when thinking about the findings of Chien et al. (2014) it is important to note the very small sample size (N= 13 in both groups) and the fact that the ASD group was matched for age but not ability to the TD group. Most importantly, all studies to date have looked at the ORA in ASD in single cultures (predominantly Chinese participants) and the literature currently lacks a bi-directional study of the ORA.

The Current Study

The aim of this study was to investigate how experience shapes face perception in ASD by studying the ORA cross-culturally. Crucially, this is the first simultaneous cross-cultural study of children with autism for the ORA to the authors' knowledge. We compared children with and without autism in both the UK and Japan using a face recognition task adapted from Chien et al. (2014) [two-alternative forced-choice task, Asian and Caucasian faces, 4 conditions of difficulty]. It was expected that TD children in the UK and Japan would show an ORA highlighting how cultural experience with native faces shapes face perception. Leading from the findings of Chien et al. (2014) it was expected that the ORA would be seen most clearly in the identity change condition (the easiest condition where children should show the greatest level of expertise and accuracy). Based on theories of reduced perceptual experience, it was predicted that the ASD groups in the UK and Japan (matched to TD children for age and cognitive ability) would showed a reduced ORA, and show particular difficulties on the conditions where manipulations are made to the eye regions of own and other-race faces. Given the findings of Wilson et al. (2011) we were also mindful to explore possible heterogeneity of the ORA in children with ASD. This timely investigation provides the first evidence of how cultural experience shapes face perception in ASD simultaneously from two different cultural groups.

Method

Participants

124 children were recruited to participate in the study, including 56 children from the UK and 68 from Japan. All participants had normal or corrected-to-normal vision. Comparison of the participants from both samples (Japan vs. UK) showed that the Japanese sample were significantly older than the UK sample (on average 9.91 years vs. 9.07), t (122) = -2.5, p = .014, d = -.71, and they also had a significantly higher average RCPM score (30.43 vs. 28.43), t(104.14) = 2.36, p = .020, d = .43. It was not possible to compare verbal

ability given the different measures used. Given the differences between the samples[2], age-matched subgroups (total n=84) were constructed from the larger full sample (total n=124) and the analysis focused on these groups[3]. It was possible to match 16 children with ASD (mean 121.4m, SD 21.4) and 26 TD children (mean 117.4m, SD 24) to 16 children with ASD from Japan (mean 121.9m, SD 19.9) and 26 TD children (mean 117.6m, SD 23.8). There were no differences in age between the ASD children in the UK and Japan, $t(30) = .077$, $p = .939$, $d = .03$, or between the TD children in the UK and Japan, $t(50) = -.023$, $p = .982$, $d = .008$. Additionally, none of the groups differed on Ravens scores (all comparisons UK/Japan, ASD/TD: t's < .347, p's >.731). Table 1 presents the data on the participant characteristics.

In the UK sample, all children were recruited through local schools and advertisements. The age range for the UK sample was 6.8 to 13.4 years. The UK-ASD group consisted of 16 children (2 female, 14 male), with a mean age of 10.1 years (SD 1.7). Children with autism had all been previously diagnosed by experienced clinicians according to the DSM-IV criteria (American Psychiatric Association, 1994), and they all had a full statement of special educational needs and/or an education health and care plan (EHCP). The TD group consisted of 26 children (15 female, 11 male), with a mean age of 9.8 years (SD 2.0). Verbal ability was assessed using the British Picture Vocabulary Scale III (BPVS; Dunn & Dunn, 2009), and non-verbal ability was assessed with Ravens Colour Progressive Matrices (RCPM; Raven, Raven & Court, 1998b). The ASD group had a mean standardised BPVS score of 99.8 and a mean RCPM score 29.8. The TD group had a mean standardised BPVS score of 105.6, and a mean RCPM score of 29.6. There were no significant differences

[2] Age and RCPM were found to correlate with accuracy in both samples, UK age, r(52) = .254, p = .03, RCPM, r(52) = .187, p = .09; Japan, Age, r(68) = .329, p = .003, RCPM, r(68) = .275, p = .01.
[3] Analysis of the data with the larger samples is provided in the Supporting Information, although these are within-cultural-sample analyses given the differences in age and Ravens between the UK and Japan groups.

between ASD and TD groups for age, *t*(40) = .529, *p* = .600, *d* = .017, standardised BPVS scores, *t*(54) = -1.517, *p* = .137, *d* = .49, or RCPM score, *t*(40) = .122, *p* = .904, *d* = .039.

Table 1: A comparison of the descriptive statistics of the ASD and TD groups of the UK and Japanese samples

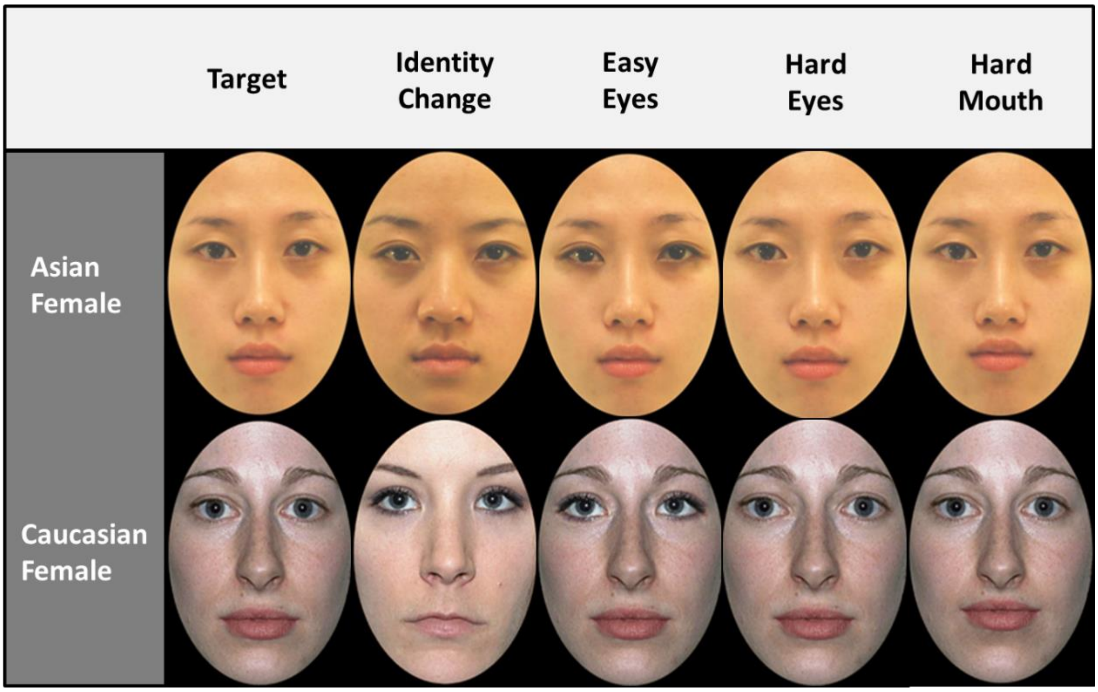
		ASD	**TD**
Gender (M/F)	**UK**	17/4	21/14
	Japan	21/8	22/17
Age (Years) M (SD)	**UK**	9.10 (2.19)	9.06 (1.77)
	Japan	9.90 (1.52)	9.92 (2.03)
Peabody M (SD)	**UK**	N/A	N/A
	Japan	10.83 (1.38)	10.65 (1.64)
Ravens M (SD)	**UK**	28.14 (6.13)	28.57 (4.63)
	Japan	30.72 (3.73)	30.21 (4.44)
BPVS M (SD)	**UK**	100.48 (13.34)	104.51 (12.78)
	Japan	N/A	N/A

The age range for the Japanese sample was 6.6 to 13.3 years. In the Japanese sample, individuals with ASD were recruited through a medical university in the Tochigi area. The diagnosis of the participants with ASD was based on the DSM-IV criteria (APA, 1994) and was confirmed by trained paediatric neurologists. Typically developing children were recruited from elementary, junior high, and high schools near in the Tochigi area. The ASD group consisted of 16 children (1 female, 15 male), with a mean age of 10.1 years (SD 1.63). The TD group was made up of 26 children (9 female, 17 male), with a mean age of 9.8 years (SD 1.98). Verbal ability was assessed using the Japanese version of the Peabody Vocabulary Test, referred to as the Picture Vocabulary Test (PVT; Ueno, Utsuo & Iinaga, 1991), and non-verbal ability was assessed using the RCPM (Raven et al., 1998b). The ASD group had a mean PVT scaled score of 10.5, and a mean RCPM of 30.3. The TD group had a mean PVT score of 10.1, and a mean RCPM of 29.8. There were no significant differences between the TD and ASD groups for age, $t(40) = .605$, $p = .549$, $d = .19$, PVT scores, $t(40) = .829$, $p = .412$, $d = .266$, or Ravens, $t(40) = .347$, $p = .713$, $d = .11$.

Apparatus and Stimuli

The stimuli were colour images of Caucasian and Asian male and female adult faces, cropped and framed in an oval-shape window to remove the background and external cues, and mounted on a black background. The images were from a stimuli set used in several published papers on the ORA in infants, children and adults (Chien et al., 2014, Chien, Wang & Huang, 2016, Chien et al., 2018). The Asian faces were selected from the Taiwanese Facial Expression Image Database (TFEID; Chen & Yeh, 2007), while the Caucasian faces were from the NimStim Face Stimulus Set (Tottenham et al., 2009). The selected face images were with a frontal pose, neutral expression, and with no glasses, or hair covering the forehead. The skin tones of individual faces within the same race were rendered equal to reduce differences in color and luminance by PhotoImpact 10 software (Ulead System, Taipei). For further details on stimuli, see Chien et al., 2014 and Hsu & Chien, 2011. There were four target stimuli: male Caucasian, female Caucasian, male Asian and female Asian. In each recognition trial, target faces were presented alongside a foil face from one of four difficulty conditions. These were: identity change (different face to target), easy eyes (same target face, different eyes photoshopped in), hard eyes (same target face, eye spacing increased), and hard mouth (same target face, nose-mouth spacing increased). The locations of the target and foil were counterbalanced. An example of the Caucasian and Asian female faces is provided in Figure 1. In the recognition trials, the dimensions for the face stimuli 8.5cm wide and 11.5cm long. At a viewing distance of approximately 60 cm (as per our testing protocol) this equated to 10.5° by 8° of visual angle.

Figure 1: Examples of Caucasian and Asian female faces showing the target, and the foil shown in each difficulty condition.



The experimental program was compiled with E-prime 2.0 software (Psychology Software Tools, Inc., PA, USA), and run on a laptop computer (UK: Lenovo N500; Japan: HP Pavilion Desktop, h8-1060jp) which was connected to a monitor for display (UK: 22 inch; Japan: 23-inch). There were 64 trials in total, [4 identities (2 Asian, 2 Caucasian) X 4 conditions (IC, EE, HE, HM) X 2 locations for the correct answer (left, right) X 2 repetitions]. 32 trials were presented, followed by a short break during which the participant was required to stay seated, followed by the final 32 trials. Prior to testing each child completed a 4 trial practice session. These mirrored the identity change condition, but were not included in the formal test stimulus set.

Procedure

Children in the UK and in Japan were tested in a quiet room, free from distractions in either their school or a lab. The face recognition task was completed first, followed by cognitive measures of verbal and non-verbal ability. Faces were presented using a sequential two-alternative-forced-choice discrimination trial. Each child was positioned in front of the monitor. The trial began with a fixation cross, followed by presentation of the target face. After 3 seconds, the target face disappeared, followed by a blank screen for 1 second, and then the test trial which displayed two faces, one on the left and one on the right. Children were instructed to indicate which face was the same as the one they had just seen, by pressing either the left or right key on a controller. The two faces remained on screen until a response was made, and no time limit was enforced. Feedback was not given throughout the trial, although the children were positively encouraged during the task.

Ethical approval was obtained from the local ethical committees in the UK and Japan relevant to each institution. Informed written consent was obtained from the parents of all children and assent was obtained from all children before taking part in the study.

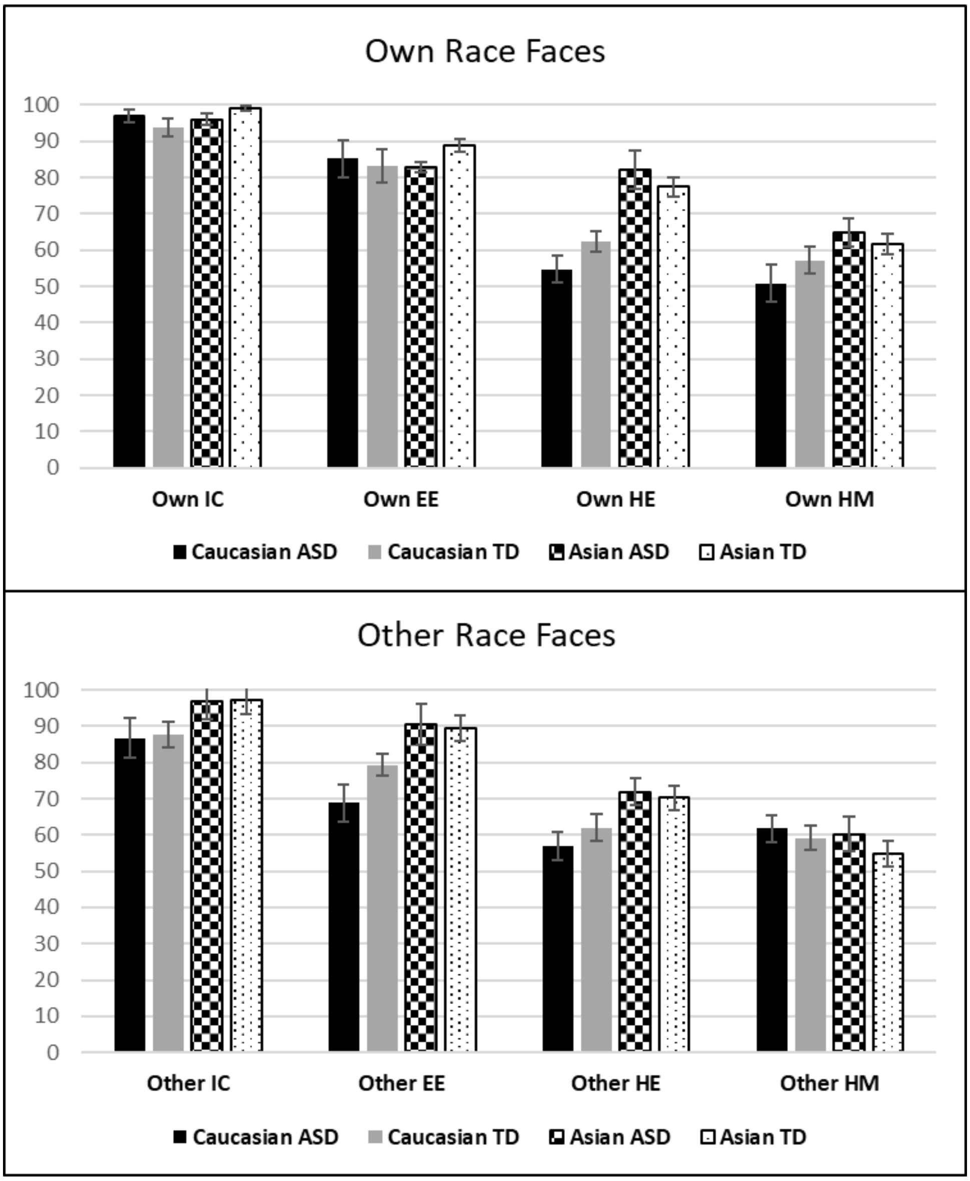
Results

Do children with autism show an own-race advantage?

Figure 2 presents the recognition accuracy data for all groups (ASD, TD: UK, Japan). To examine face recognition performance, a four-way mixed ANOVA was carried out using within subject factors face type (own and other) and difficulty level (IC, EE, HE, HM), and the between subjects factors of developmental group (ASD, TD) and cultural group (Japan, UK). There were main effects of face type, $F(1, 80) = 8.458$, $p = .005$, $\eta^2_p = .09$, difficulty, $F(3, 240) = 148.209$, $p <.001$, $\eta^2_p = .649$, and culture, $F(1, 80) = 15.00$, $p <.001$, $\eta^2_p = .158$. There was a two-way interaction effect between difficulty and culture, $F(3, 240) = 4.658$, $p = .003$, $\eta^2_p = .05$. Finally, there was a three-way interaction between face type, difficulty and culture, $F(3, 240) = 10.174$, $p <.001$, $\eta^2_p = .113$. The four way interaction between face type, difficulty, culture and group did not reach statistical significance, $F(3, 240) = 2.402$, $p = .068$, $\eta^2_p = .029$. All other main effects and interactions did not reach statistical significance (all *F*'s < 1.044; all *p*'s > .361).

The main effect of face type indicated that on average children were better at recognising own race (mean 77.3%, SD, 10.6) compared to other race faces (mean, 74.3%, SD, 12). The main effect of culture was driven by better accuracy overall in the Japan group (mean 79.9 %, SD 8.99) compared to the UK group (mean 71.7 %, SD, 10.5). To understand the three-way interaction (which subsumed the two-way interaction between difficulty and culture), we ran separate two-way ANOVAs with factors face type (own, other) and difficulty (IC, EE, HE, HM) for each cultural group to understand what was happening within each sample, and then we ran two-way ANOVAs to compare the cultural samples at each difficulty level, separately for each face type (own, other).

Figure 2: Accuracy for face recognition (Own Race Faces on top; Other Race Faces on bottom) in each difficulty level for ASD and TD groups in Japan and the UK. Error bars represent +/- 1 SE of the mean.



Comparisons within Japan and UK samples. UK sample: A two-way ANOVA with factors face type (own, other) and difficulty (IC, EE, HE, HM) revealed a main effect of face type, $F(1, 41) = 4.015$, $p = .052$, $\eta^2_p = .089$, a main effect of difficulty, $F(3, 123) = 71.746$, $p < .001$, $\eta^2_p = .636$, and an interaction between face type and difficulty, $F(3, 123) = 5.510$, $p = .001$, $\eta^2_p = .118$. Children from the UK were more accurate at recognising own (mean 73.2%, SD 10.1) compared to other (mean 70.16%, SD 12.5) race faces. Pairwise comparisons showed that performance across the difficulty levels was as expected, with IC being the easiest (mean 94%; significantly different to all other levels, all *p*'s <.001); followed by EE (85.7%; significantly different to all other levels, all *p*'s <.001); followed by HE (67.8%) which was easier than HM (61.1%) although the difference between these levels was not significant ($p = .226$).

To unpick the two-way interaction paired t-tests were used to compare each difficulty level within each face category (Bonferroni correction applied for multiple comparisons: α .05 / 6 = .008), and then further paired t-tests were used to compare each difficulty level between the face race categories (Bonferroni correction applied for multiple comparisons: α .05 / 4 = .0125). Within the own race face category, all difficulty levels were significantly different from each other (IC > EE > HE; all *t*'s >6.232, all *p*'s <.001), except for the HE and HM which were equally difficult, t (41) = 1.376, p = 176, d = .25. Within the other race face category, the pattern was the same (IC > EE > HE; all *t*'s >3.550, all *p*'s <.002) and HE and HM were again not significantly different from each other, t (41) = .001, p = 1.000, d = <.001. When comparing own and other race face recognition for each difficulty level, significantly better accuracy for own race faces was observed for IC, t (41) = -3.499, p <.001, d= .44, EE, t (41) = -3.669, p <.001, d = .49, but not for HE, t (41) = .167, p = .868, d =.03 or HM, t (41) = 1.851, p = .071, d = .31 .

In sum, for the UK group HE and HM were the most difficult conditions and although own races faces were easier to recognise than other race faces, this advantage was not present in the most difficult (HE and HM) conditions for children in the UK.

Japan sample: A two-way ANOVA with factors face type (own, other) and difficulty (IC, EE, HE, HM) revealed a main effect of face type, $F(1, 41) = 5.620$, $p = .023$, $\eta^2_p = .121$, a main effect of difficulty, $F(3, 123) = 95.385$, $p < .001$, $\eta^2_p = .699$, and an interaction between face type and difficulty, $F(3, 123) = 3.609$, $p = .015$, $\eta^2_p = .081$. Children from Japan were more accurate at recognising own (mean 81.3%, SD 9.7) compared to other (mean 78.57%, SD 10.1) race faces. Pairwise comparisons showed that performance across the difficulty levels was as expected, with IC being the easiest (mean 94.7%; significantly different to all other levels, all *p*'s <.001); followed by EE (84.3%; significantly different to all other levels, all *p*'s <.001); followed by HE (69.6%) which was significantly easier than HM (58.9%; $p = .006$).

To unpick the two-way interaction paired t-tests were used to compare each difficulty level within each face category (Bonferroni correction applied for multiple comparisons: α .05 / 6 = .008), and then further paired t-tests were used to compare each difficulty level between the face race categories (Bonferroni correction applied for multiple comparisons: α .05 / 4 = .0125). Within the own race face category, IC was easier than EE, $t(41) = 4.500$, $p < .001$, $d = .9$, EE easier than HE, $t(41) = 2.546$, $p = .015$, $d = .40$, and HE was easier than HM, $t(41) = 4.394$, $p < .001$, $d = .90$ (all other comparisons *t*'s < 23.89, *p*'s <.001). Within the other race face category, all levels were significantly different from each other (IC > EE > HE > HM; all *t*'s > 3.344, *p*'s < .003). When comparing own and other race face recognition for each difficulty level, significantly better accuracy for own race faces was observed for HE, *t* (41) = -2.895, *p* =.006, *d* = .42, but no differences were found for the other difficulty levels: IC, *t* (41) = -.650, *p* =.519, *d* = .13, EE, *t* (41) = 1.567, *p* = .125, *d* =.19, or HM, *t* (41) = -1.580, *p* = .122, *d* =.34.

In sum, recognition accuracy decreased across each level showing that children in the Japan group found each difficulty level harder than the last, for both own and other races faces. Although the overall accuracy performance indicated that own races faces were easier to recognise than other races faces for the Japan group, this was driven by better recognition for own race faces in the HE condition only.

Comparisons between Japan and UK samples. A two-way ANOVA with factors culture (UK, Japan) and difficulty level (IC, EE, HE, HM) for recognition accuracy of own race faces revealed a main effect of culture *F*(1, 82) = 16.482, *p* <.001, η^2_p = .167, a main effect of difficulty, *F*(3, 246) = 53.392, *p* <.001, η^2_p = .394 and an interaction between the two factors, *F*(3, 246) = 5.604, *p* =.001, η^2_p = .064. To unpick the interaction, independent t tests were conducted to compare the cultural groups for each difficulty level (Bonferroni correction applied for multiple comparisons: α .05 / 4 = .0125).

There was no difference between the children in the UK and Japan for the own race IC condition, *t* (82) = -1.604, *p* =.113, *d* = .34, the EE condition, *t* (82) = -.782, *p* =.437, *d* = .17, the HM condition, *t* (82) = -2.172, *p* =.033, *d* =.47, whereas they were significantly different for HE, *t* (82) = -4.577, *p* =.001, *d* =.1.003 The children from Japan had higher accuracy for faces (mean 79.1%, SD 19.4) in the own race HE condition compared to children in the UK (mean 59.5, SD, 19.8).

A two-way ANOVA with factors culture (UK, Japan) and difficulty level (IC, EE, HE, HM) for recognition accuracy of other race faces revealed a main effect of culture, $F(1, 82) = 10.219$, $p = .002$, $\eta^2_p = .111$, a main effect of difficulty, $F(3, 246) = 85.639$, $p < .001$, $\eta^2_p = .511$ and an interaction between the two factors, $F(3, 246) = 5.434$, $p = .001$, $\eta^2_p = .062$. To unpick the interaction, independent t tests were conducted to compare the cultural groups for each difficulty level (Bonferroni correction applied for multiple comparisons: α .05 / 4 = .0125). The cultural groups were significantly different for the IC, $t(82) = -2.730$, $p = .008$, $d = .59$, EE, $t(82) = -3.834$, $p < .001$, $d = .83$, and HE conditions, $t(82) = -2.638$, $p = .010$, $d = .57$, but not for the HM condition, $t(82) = .868$, $p = .388$, $d = .18$. For those conditions that were significantly different between the groups, children in Japan were more accurate (IC 97%; EE 89.8%; HE 70.8%) than children in the UK (IC 87%; EE 75.2%; HE 60.1%).

In sum, when comparing children between each cultural group, children in the Japan group had higher accuracy compared to children in the UK group for own race faces in the HE condition, and higher accuracy for other race faces in the IC, EE and HE conditions.

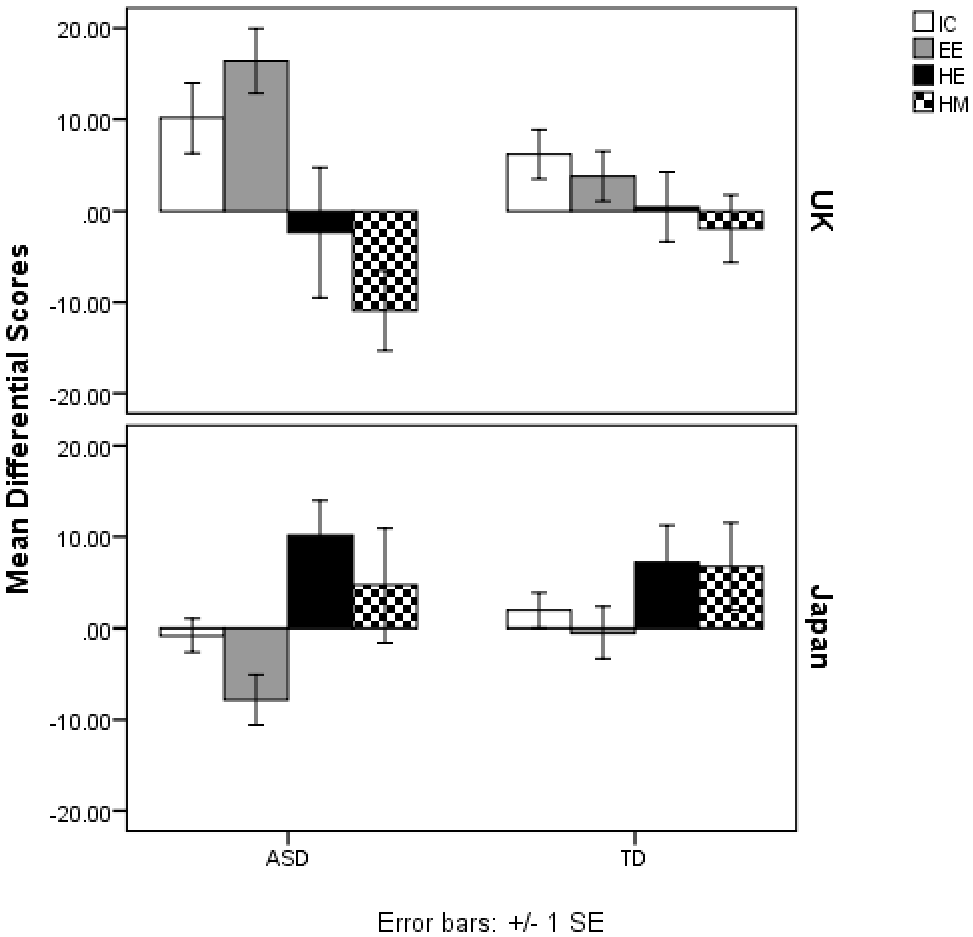
Summary for age-matched mixed ANOVA. When analysing the data to include both between group factors – developmental group *and* cultural group - there were no significant effects attributable to developmental group, but there were to cultural group membership. Even though the children in both cultural groups were matched for age and non-verbal reasoning ability, the difficulty level of each condition affected group performance differently. For example, children in the UK found the HE and HM equally difficult and they did not show an ORA in the HE and HM conditions, whereas children in Japan found each difficulty level harder than the last, and only showed an ORA for the HE condition. Indeed, it was the HE condition that revealed the significant differences in accuracy when the groups were compared directly, with children in Japan showing higher accuracy in HE for own *and* other race faces. Additionally, children in Japan had higher accuracy than the UK group for other race faces in the IC and EE conditions.

Differential scores. The ORA was further probed by analysing the differential scores, which not only allow for a more detailed look at the ORA but also an opportunity to explore heterogeneity. Differential scores are made by subtracting total other-race performance for each condition from own-race performance. Thus, a positive score indicates an ORA, whilst a negative score indicates an other-race advantage. Figure 3 shows the differential scores for all of the groups. Firstly, one samples t tests were used to test the presence of the ORA for each group (UK TD, UK ASD, Japan TD, Japan ASD) in each difficulty level. For TD children in the UK, there was a significant ORA in the IC condition, $t\ (25) = 2.308$, $p = .03$, but not in any other conditions (all *t*'s < 1.397, all *p*'s >.175). For ASD children in the UK, there was a significant ORA in the IC, $t\ (15) = 2.657$, $p = .018$ and EE , $t\ (15) = 4.612$, $p <.001$ conditions but a significant *other-race advantage* in the HM condition, , $t\ (15) = -2.485$, $p = .025$. There was no significant effect in the HE condition, $t\ (15) = -.328$, $p = .747$.

For TD children in Japan there was no significant ORA found in any condition (all *t*'s < 1.779, all *p*'s >.087). For ASD children in Japan, there was a significant ORA in the HE condition, $t\ (15) = 2.657$, $p = .018$, but a significant *other-race advantage* in the EE condition, $t\ (15) = -2.825$, $p = .013$ (for IC and HM, t's <.752, p's >.464)[4].

[4] Bivariate correlations were conducted between age and differential scores for each condition for the entire sample (N = 84; Bonferoni corrected for multiple tests: α .05 / 4 = .0125) and no correlations were significant (all *r*'s <.228, all *p*'s <.037). Furthermore, we checked to see if age or Ravens were correlated with differential scores for all ASD children or all TD children. None of these correlations were significant (all *r*'s < .308, *p*'s >.087)

Figure 3: Mean differential scores for the ASD and TD groups in the UK and Japan.



In order to explore how development *and* culture may impact upon the ORA, a three-way mixed ANOVA with factors developmental group (ASD, TD), cultural group (UK, Japan) and difficulty level (IC, EE, HE, HM) was conducted on the differential scores in the age-matched samples. There was no main effect of developmental group, $F(1, 80) = .091$, $p =.764$, $\eta^2_p = .001$, cultural group, $F(1, 80) <.001$, $p =.984$, $\eta^2_p <.001$, difficulty level, $F(3, 240) = 1.109$, $p =.346$, $\eta^2_p = .014$, nor was there an interaction between developmental group and difficultly level, $F(3, 240) = .740$, $p =.529$, $\eta^2_p = .009$. There was also no three-way interaction effect between developmental group, cultural group and difficulty level, $F(3, 240)$

= 2.402, p =.068, η^2_p = .029. There was a significant two-way interaction between cultural group and difficulty, $F(3, 240) = 10.174$, p <.001, η^2_p = .113. In order to unpick this interaction, paired t tests comparing each condition were conducted separately for each cultural group, followed by independent t tests comparing the cultural groups for each condition.

Paired t tests (Bonferroni correction applied for multiple comparisons: α .05 / 6 = .008) showed that for children in the UK, differential scores were significantly higher in IC compared to HM, t (41) = 3.675, p = .001, d = .78, and in EE compared to HM t (41) = 3.438, p = .001, d = .81, but no other comparison were significant (all t's < 2.308; all p's >.026). Paired t tests (Bonferroni correction applied for multiple comparisons: α .05 / 6 = .008) showed that for children in Japan, differential scores were significantly higher in the HE compared to EE condition, t (41) = -3.729, p = .001, d = .71, but no other comparison were significant (all t's < -2.199; all p's >.034).

Independent t tests (Bonferroni correction applied for multiple comparisons: α .05 / 4 = .0125) comparing the UK and Japanese groups for each level of difficulty revealed that the UK group had significantly higher differential scores compared to the Japan group in the IC, t (82) = 2.629, p = .01, d = .57, and EE conditions, , t (82) = 3.784, p <.001, d = .82. The comparisons between the groups for the HE and HM conditions were not significant [HE: , t (82) = -1.949, p = .055, d = .42; HM: , t (82) = -2.380, p = .020, d = .52].

In sum, when considering the data in terms of *bias* for own-race faces over other race faces, different patterns emerged between the cultural groups and also between the developmental groups. The ORA was clearly more evident in the UK sample compared to the

Japan sample (especially in the IC and EE conditions). Although it appeared that the Japan sample had an ORA in the HE condition, this was in fact only for the Japan ASD group. The Japan TD did not show an ORA in any condition. Furthermore, although the UK ASD group did show an ORA in the IC and EE conditions, similar to the UK TD group, they also showed an *other-race advantage* in the HM condition. Indeed, while the Japan ASD group showed an ORA in the HE condition, they too also showed an *other-race advantage* in the EE condition.

Discussion

The aim of this study was to investigate how culture shapes face recognition in ASD by testing the ORA cross-culturally in Japan and the UK. It was hypothesised that based on atypical experience with faces in ASD (Dawson et al., 2004; Chevalier et al., 2012), there would be a reduced and atypical ORA (Chien et al., 2014), and that accuracy would be lower for children with autism particularly in the conditions that relied on information from the eyes (Chien et al., 2014). Contrary to expectation, we found that children with ASD in the UK and in Japan showed typical face recognition ability and did not have particular difficulties in conditions where recognition relied on the eye region (EE, HE). They also did not show particular advantages for recognition dependent on the mouth region, in contrast to Chien et al. (2014). Furthermore, there was no clear evidence of a reduced ORA in ASD.

When testing for the effect of developmental group *and* cultural group simultaneously, there were no developmental group effects but there were cultural group effects. Although on average children were better at recognising own compared to other-race faces, and performance in the difficulty conditions was as expected (IC > EE > HE > HM;

note for children in the UK HE = HM), children in Japan had higher accuracy than children in the UK. The results of the three-way interaction between culture, face type and difficulty showed that this was specific to certain conditions - children in the Japan group had higher accuracy compared to children in the UK group for own race faces in the HE condition, and higher accuracy for other race faces in the IC, EE and HE conditions. Therefore, the HE condition was a particular strength for children from Japan compared to children in the UK.

Finally, analysis of differential scores showed the conditions in which there was a bias for own compared to other-race faces. The ORA was clearly more evident in the UK sample compared to the Japan sample (especially in the IC and EE conditions). Although it appeared that the Japan sample had an ORA in the HE condition, this was in fact only for the Japan ASD group. The Japan TD group did not show an ORA in any condition. An interesting aspect to these data was that there was considerable heterogeneity within-groups in terms of the presence of the ORA (see Figure 3). It is worth noting that the heterogeneity appeared to be larger in the autism groups in both Japan and the UK, and while the trends for the TD groups were in the direction of an ORA, there were clear trends in the opposite direction for the autism groups (for an other-race advantage). Indeed, while both ASD groups showed an ORA in some conditions, they also showed *an other-race advantage* in other conditions.

The ORA and ASD

The present study, for the first time, reported the data from separate cultural groups of TD and ASD participants, providing evidence that experience of faces from one's own cultural group shapes face perception in children with ASD. Importantly, for the first time we report data from separate cultural groups of TD and ASD groups studied simultaneously. Our findings with UK children are most similar to Wilson et al. (2011), as both studies observed and ORA in both TD children and children with ASD. Our findings with children from Japan did not support Yi et al. (2016) or Chein et al. (2014) both because our Japanese children with ASD did show an ORA and because our TD children from Japan did not show an ORA. This is particularly interesting given that we used the same task as Chien et al. (2014) using (albeit with Asian/Caucasian faces, as opposed to Asian/African faces). Chien et al. (2014) reported a lack of an ORA in ASD, comparing thirteen age-matched children with ASD to TD children. Looking more closely at their differential percent scores, it seems that the small sample size may be an important consideration in relation to the ORA, as it was only observed in TD in the IC change condition, and in two further conditions (EE and HE) the ASD group appeared to be showing positive trends in favour of an ORA, albeit with considerable variability. It may well be the case that increased sample size and power in Chien et al. (2014) would have led to different interpretations of the results. Indeed, with a larger sample size (ASD N = 27) variability and heterogeneity was highlighted by Wilson et al. (2011) as when looking in more detail at within-group performance, despite finding a typical ORA in ASD overall it was reported that a sub-group of children in the ASD group who had face recognition scores below expectations for their age did show an atypically reduced ORA. Therefore, variability and heterogeneity is important for understanding the ORA in ASD.

Looking in more detail at within-group performance using differential percent scores in the current study, it is possible to note some interesting trends and heterogeneity. For example, the Japan TD group do not show a clear ORA. However, the two typically developing groups in Japan and the UK show a more consistent pattern with each other than the two ASD groups. The two typically developing groups show a pattern that is more positively biased in all conditions whereas the two ASD groups show a much larger range of scores and clear negative biases in some conditions and *an other-race advantage*. Therefore, although the results provide evidence for an ORA in ASD in some conditions, they also show an *other-race advantage* and indications of significant variability are worth probing in more depth in future studies. Variance in differential scores was not related to either age or non-verbal ability and it would be interesting, and important, to look at other factors that might predict this heterogeneity in future work.

Cross-cultural differences

The age-matched analysis allowed us to make direct cross-cultural comparisons and it was found that Japanese children had higher accuracy compared to children in the UK (by 8.2% on average) despite being matched for age and non-verbal ability. More specifically, performance was better in the own-race HE condition as well as the other-race HE, IC and EE conditions. It is not clear exactly why this difference was observed, although it is possible that is related to cross-cultural differences in face scanning patterns. Several studies have shown that Eastern observers fixate the eyes more than other regions of the face, compared to Western viewers who scan more features, particularly the mouth (Jack, Blais, Scheepers, Schyns & Caldara, 2009; Senju, Vernetti, Kikuchi, Akechi & Hasegawa, 2013). It has been suggested that such strategies explain why Eastern observers perform relatively poorly compared to Western participants on some tests of emotion recognition (Jack et al., 2009). However, it is possible that such cross-cultural differences in scanning provided an advantage to the children from Japan in this study. This is an interesting finding that should be followed-up in future studies.

Additionally, although the UK TD group showed an ORA in the IC condition, the Japan TD group did not show a reliable ORA. These children were of an age where it would be expected that an ORA would be observed (Chance et al., 1982; Chien et al., 2018; Goodman et al., 2007) and Chien et al. (2014) reported an ORA in an Asian TD sample with a very similar age range using very similar stimuli (although the ‘other-race’ category were African as opposed to Caucasian faces). Therefore this does not appear to be a systematic cultural difference or a particular effect with these stimuli. However, we are not the first study to find that an Asian TD group has not shown the expected ORA, as outlined in the introduction (Yi et al., 2015). Indeed, the effect observed in Chien et al. (2014) was relatively small (ORA observed in one out of four conditions). This may reflect the possibility that the ORA is context-dependent (e.g. difficulty of the recognition task) or that it is not universally observed (e.g. not all other-races faces are equally unfamiliar). Future work should probe this further and explore potential factors that may help to explain individual differences in the ORA in TD and ASD groups.

Implications for understanding ASD

The rationale in the current study for predicting a reduced ORA in ASD was developed from the literature showing reduced and atypical experience with faces in ASD (Dawson et al., 2004; Hanley et al., 2014; Jones et al., 2008; Klin, et al., 2002). Although there is clear evidence that experience with faces is atypical in ASD, it is important to consider why the ORA may develop typically. One possible explanation may be linked to attention to faces in early infancy. We know that in typical development, the origins of the ORA are established early, and that over the first year of life infants preferentially attend to the faces in their immediate environment and develop more accurate recognition for own-race faces (Kelly et al., 2007a). Thinking about early experience with faces in ASD, emerging evidence involving infants who go on to receive a diagnosis of an ASD indicates that divergence from typical social attention may only be clear towards the end of the first year of life and there may be vast heterogeneity. In their review, Webb, Neuhaus and Faja (2017) reported that attention to faces, face perception, and face learning memory in ASD diverge from typical development somewhere in the second half of the first year of life, before a more apparent delay and atypicality is seen in toddlerhood, though again with vast heterogeneity. Related to this, eye-tracking work by Jones and Klin (2013) shows that in terms of attention to others' eyes, children who later go on to be diagnosed with ASD show typical patterns of attention to the eyes at birth. This attention pattern begins to diverge after about 2 months of age, and eye fixations gradually decline (increase in face scanning atypicality) over time from 2 to 12 months. The important consideration here is that attention in ASD appears typical at and shortly after birth, and then increases in atypicality over time. It may be that experience of faces in the first year of life, linking to the evidence on the typical development of the ORA during infancy, is sufficient for a 'typical' ORA to develop in at least some children with ASD. Of course the children who participated in the current study are much older than this and therefore we can only hypothesise about how earlier experiences may have impacted upon the pattern we see here in later childhood.

It may also be that beyond the first year of life, there is significant heterogeneity in autism in terms of experience (e.g. social attention) which impacts upon the development of face expertise (e.g. face recognition), which in turn impacts on the ORA. Interestingly, in two of the three studies that have found a reduced ORA in ASD, face recognition atypicalities were also reported. Hadad et al. (2019) showed that the ORA was reduced in a group of autistic adults who also showed poorer recognition and a reduced inversion effect for own-race faces in comparison to TD adults. Wilson et al. (2011) found that although they observed a typical ORA for children with autism at the group level, there was a sub-group of children with autism with age inappropriate face recognition who did not show a typical ORA. Therefore, future work is needed with much larger samples to probe individual differences in the ORA in relation to social attention and face recognition ability, particularly in children when these skills are still developing.

An alternative, though not mutually exclusive, explanation is that atypical experience with faces in ASD (certainly from the first year onwards, Webb et al., 2017; Jones & Klin, 2013) has a differential impact on the variety of different face processing skills. Not all elements of face perception require the same skills or degree of expertise. For example, whilst deciphering complex socio-cognitive information from faces, such as mental state recognition, can be particularly problematic (Celani et al., 1999; Gross, 2004, Nuske et al., 2013), individuals with ASD have less difficulty with face recognition (Walsh, Creighton & Rutherford, 2016). Arguably, emotion and mental state recognition are more complex than face recognition; they require the ability to extract information from throughout the face (e.g. facial configural information) and a higher level socio-cognitive skills for processing and interpretation. In contrast, face recognition arguably involves more visual perceptual skills. Walsh et al. (2016) suggest this to be the case, reporting that participants with ASD showed emotion recognition deficits alongside relatively more proficient face identification ability (at typical levels). The task used in the current study relied predominantly on visual perceptual skills. Children were required to identify which image (from two) matched the one they had previously seen. Therefore, it may have captured image-matching skills as opposed to more sophisticated face processing skills, such as those involved in learning face identities in everyday life (although it is highly relevant that the task used here matched that used in previously published studies). Andrews, Burton, Schweinberger and Wiese (2017) report that learning instances of a single face image (as in this study) is insufficient to recognise other instances of the same face. In other words, the face itself has not been learned, but only one visual image of that face. In order to more fully understand the way culture shapes face perception in typical and atypical development, future work on the ORA could involve tasks that probe how faces are learned in everyday life.

Limitations and Conclusions

It is worth highlighting several considerations which should be addressed in future research. The face recognition task used in this study was chosen for several reasons, including the fact that it had been used in a similar study in the area (Chien et al., 2014), thus allowing us to probe issues around inconsistency in the literature. It also offered an opportunity to explore face recognition ability in ASD linked to eye and mouth region reliance. However, the measure the task yielded was accuracy, and it would be useful in future studies to measure attention while participants complete such tasks. The use of eye-tracking techniques would allow the possibly to directly link attention to performance (Chita-Tegmark, 2016, Corden et al., 2008) and thus reveal whether all groups perform the task in similar ways. It would also allow further analysis of the hypothesis linking reduced attention to faces and an atypical ORA. Finally, the addition of eye-tracking techniques with a cross-cultural comparison would allow a different way to probe how culture shapes perception by investigating whether children with ASD show the same cultural face scanning patterns as their typically developing counterparts, and whether cross-cultural differences in face scanning link to recognition accuracy (Yi et al., 2015).

Although the sample sizes in the current study were larger than other studies in the area (Chien et al., 2014; Yi et al., 2015; Yi et al., 2016), future work with larger samples would allow more in-depth analysis of within-groups heterogeneity. This may be particularly interesting and important for understanding face perception in ASD, where this skill may show considerable within-group variability (Wilson et al., 2011). Future work could also probe heterogeneity further by looking at links to individual differences such as with autism symptom measures.

In conclusion, the current study is the first to investigate the ORA in ASD cross-culturally. Our findings do not support the hypothesis that atypical experience with faces leads to a reduced ORA for children with autism. Instead, our results emphasise a really variable profile in ASD where both and ORA and an other-race advantage can be seen at the same time. In doing so, this study significantly adds to the literature and provides an insight into some of the inconsistencies previously seen. The findings also show interesting cross-cultural differences in performance which emphasise why we cannot assume that research findings on the ORA apply universally. Future research should focus on understanding the factors that are associated with heterogeneity in face recognition in ASD and TD, as these are skills are crucial for guiding interpersonal communication.

References

Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T. & Hietanen, J.K. (2013). Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings. *PLoS One*, 8(3): e59312. doi:10.1371/journal.pone.0059312

Andrews, S., Burton, M. A., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *The Quarterly Journal of Experimental Psychology, 70(8),* 1620-1632. doi: 10.1080/17470218.2016.1195851

Anzures, G., Wheeler, A., Quinn, P. C., Pascalis, O., Slater, A. M., Heron-Delaney, M., ... Lee, K. (2012). Brief daily exposures to Asian females reverses perceptual narrowing for Asian faces in Caucasian infants. *Journal of Experimental Child Psychology, 112(4),* 484-495. doi: 10.1016/j.jecp.2012.04.005

Anzures, G., Kelly, D.J., Pascalis, O., Quinn, P.C., Slater, A.M., de Viviés, X. & Lee, K. (2014). Own- and Other-Race Face Identity Recognition in Children: The Effects of Pose and Feature Composition. *Developmental Psychology, 50(2):*469-81. doi: 10.1037/a0033166.

Bar-Haim, Y., Ziv, T., Lamy, D., & Hodes, R. M. (2006). Nature and nurture in own-race face processing. *Psychological Science, 17(2),* 159-163. DOI: 10.1111/j.1467-9280.2006.01679.x

Birmingham, E., Bischof, W. F. & Kingstone, A. (2008). Social attention and real world scenes: The roles of action, competition and social content. *Quarterly Journal of Experimental Psychology, 61*, 986-976

Birmingham, E. & Kingstone, A. (2009) Human social attention. *Progress in Brain Research, 176*, 309-320.

Bruce, V., Campbell, R., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S. & Wright, R. (2000). Testing face processing skills in children. *British Journal of Developmental Psychology, 18*, 319-333.

Celani, G., Battacchi, M. W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders, 29(1)*, 57-66. DOI: 10.1023/A:1025970

Chawarska, K., Klin, A., Paul, R., Volkmar, F.(2007). Autism spectrum disorder in the second year: stability and change in syndrome expression. *Journal Child Psychology and Psychiatry, 48 (2),* 128-138. DOI: 10.1111/j.1469-7610.2006.01685.x

Chance, J.E., Lockwood Turner, A. & Goldstein, A.G. (1982). Development of differential recognition for own- and other-race faces. *The Journal of Psychology, 112*, 29-37.

Chen, L.F. & Yen, Y.S. (2007). Taiwanese Facial Expression Image Database. Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan. Available from: http://bml.ym. edu.tw/~download/html

Chevalier, C., Kohls, G., Troiani, V., Brodkin, E.S.. & Schultz, R.T. (2012). The Social Motivation Theory of Autism. *Trends in Cognitive Science, 16 (4*), 231-239. doi: 10.1016/j.tics.2012.02.007

Chein, S.H-L., Tai, C-L. & Yang, S-F. (2018). The development of the own-race advantage in school-age children: A morphing face paradigm. *PLoS One*, 13(4): e0195020.

Chien, S. H-L., Wang, L-H., Chen, C. C., Chen, T. Y., & Chen, H. S. (2014). Autistic children do not exhibit an own-race advantage as compared to typically developing children. *Research in Autism Spectrum Disorders,* 8(11), 1544-1551. DOI: 10.1016/j.rasd.2014.08.005

Chien, S. H-L, Wang, J –F. & Huang, T -R (2016). Developing the Own-Race Advantage in 4-, -, and 9-Month-Old Taiwanese Infants: A Perceptual Learning Perspective. *Frontiers in Psychology*, *7*:1606. doi: 10.3389/fpsyg.2016.01606

Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities, 48,* 78-93. Doi: 10.1016/j.ridd.2015.10.011

Corden, B., Chilvers, R., & Skuse, D. (2008). Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. *Neuropsychologia, 46*, 137-147. DOI: 10.1016/j.neuropsychologia.2007.08.005

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A. & Liaw, J. (2004) Early Social Attention Impairments in Autism: Social Orienting, Joint Attention and Attention to Distress. *Developmental Psychology*, 40, 271-283. DOI: 10.1037/0012-1649.40.2.271

Dawson, G., Webb, S.J. & McPartland, L. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27 (3),* 403-424. DOI: 10.1207/s15326942dn2703_6

Dalton, K.M., Nacewicz, B.M., Johnstone, T., Schaefer, H.S., Gernsbacher, M.A., Goldsmith, H.H…..Davidson, R.J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience, 8,* 519-526. DOI: 10.1038/nn1421

De Heering, A., De Liedekerke, C., Deboni, M., & Rossion, B. (2010). The role of experience during childhood in shaping the other-race effect. *Developmental Science, 13(1),* 181-187. doi: 10.1111/j.1467-7687.2009.00876.x.

Dunn, L. M., & Dunn, D. M. (2009). *The British Picture Vocabulary Scale*. GL Assessment Limited.

Fridenson-Hayo, S., Berggren, S., Lasalle, A., Tal, S., Pigat, D., Bölte, S…..Golan, O. (2016). Basic and complex emotion recognition in children with autism: cross-cultural findings. *Molecular Autism*, 7:52. DOI: 10.1186/s13229-016-0113-9

Goodman, G.S., Hirschman, J.E., Hepps, D. & Rudy, L. (1991). Children's memory for stressful events. *Merrill-Palmer Quarterly*. 37, 109-157.

Goodman, G.S., Sayfan, L., Lee, J.S.., Sandhei, M., Walle-Olsen, A. Magnussen, S….Arredondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology, 98*, 233-242.

Gross, T. F. (2004). The perception of four basic emotions in human and nonhuman faces by children with autism and other developmental disabilities. *Journal of Abnormal Child Psychology, 32(5),* 469-480.

Hanley, M., Riby, D. M., McCormack, T., Carty, C., Coyle, L., Crozier, N., ... McPhillips, M. (2014). Attention during social interaction in children with autism: Comparison to specific language impairment, typical development, and links to social cognition. *Research in Autism Spectrum Disorders, 8(7),* 908-924. DOI: 10.1016/j.rasd.2014.03.020

Hayden, A., Bhatt, R. S., Joseph, J. E., and Tanaka, J. W. (2007). The other-race effect in infancy: evidence using a morphing technique. *Infancy* 12, 95–104. doi: 10.1111/j.1532-7078.2007.tb00235.x

Heron-Delaney, M., Anzures, G., Herbert, J. S., Quinn, P. C., Slater, A. M., Tanaka, J. W..... & Pascalis, O. (2011). Perceptual training prevents the emergence of the other race effect during infancy. *PLoS One, 6(5),* e19858. Doi: 10.1371/journal.pone.0019858

Hsu, H. Y. & Chien, S. H. L. (2011). Exploring the other-race effect in Taiwanese infants and adults. *Chinese Journal of Psychology, 53(1),* 35-57.

Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G. & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19, 1543-1548.

Johnson, M.H., Dziurawiec, S., Ellis, H. & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40,* 1-19. Doi: 10.1016/0010-0277(91)90045-6.

Jones, W., Carr, K., & Klin, A. (2008). Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers with Autism Spectrum Disorder. *Archives of General Psychiatry, 65(8),* 946-954. doi: 10.1001/archpsyc.65.8.946.

Kanwisher, N., McDermott, J. & Chun, M.M. (2007). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience, 17 (11)*, 4302-43. Doi: 10.1523/JNEUROSCI.17-11-04302.1997

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., ... Pascalis, O. (2005). Three-moth-olds, but not newborns, prefer own-race faces. *Developmental Science, 8(6),* 31-36. doi: 10.1111/j.1467-7687.2005.0434a.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K. Ge, L., & Pascalis, O. (2007a). The Other-Race Effect Develops During Infancy. Evidence of Perceptual Narrowing. *Psychological Science, 18(12)*, 1084-1089. doi: 10.1111/j.1467-9280.2007.02029.x

Kelly, D. J., Liu, S., Ge, L., Quinn, P. C., Slater, A. M., Lee, K., ... & Pascalis, O. (2007b). Cross-Race Preferences for Same-Race Faces Extend Beyond the African Versus Caucasian Contrast in 3-Month-Old Infants. *Infancy, 11(1),* 87-95. doi: 10.1080/15250000709336871

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality*?* *Journal of Experimental Child Psychology, 104(1),* 105-114. doi: 10.1016/j.jecp.2009.01.006

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59(9),* 809-816.

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R.(2008). Attention capture by faces. *Cognition, 107,* 330–342.

McKone, E., Kanwisher, N. & Duchaine, B.C. (2007) Can generic expertise explain special processing for faces? *Trends in Cognitive Science, 11,* 8-15.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7(1*), 3-35. doi: 10.1037/1076-8971.7.1.3

Mondloch, C.J. Le Grand, R. & Maurer, D. (2002) Configural face processing develops more slowly than featural face processing. *Perception*, 31 (5), 553 – 566.

Nuske, H.J., Vivanti, G. & Dissanayake, C. (2013). Are emotion impairments unique to, universal, or specific in autism spectrum disorder? A comprehensive review. *Cognition and Emotion, 27(6),* 1042-61. doi: 10.1080/02699931.2012.762900

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race face effect. *Memory & Cognition, 22(*2), 208-224. Doi: 10.3758/BF03208892

Pascalis, O., de Viviés, X., Anzures, G., Quinn, P., Slater, A.M., Tanaka, J.W. & Lee, K. (2011). Development of Face Processing. *Wiley Interdisciplinary Reviews Cognitive Science, 2(6)*: 666–675. doi:10.1002/wcs.146

Park, J., Newman, L.I. & Polk, T.A. (2009). Face Processing: The Interplay of Nature and Nurture. *The Neuroscientist*, 15, 445-449. doi: 10.1177/1073858409337742.

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: more evidence for the cross-race effect. *Journal of Applied Psychology, 88(4),* 760. Doi: 10.1037/0021-9010.88.4.760

Rapin, I., & Tuchman, R. F. (2008). Autism: definition, neurobiology, screening, diagnosis. *Pediatric Clinics of North America, 55(5),* 1129-1146. doi: 10.1016/j.pcl.2008.07.005.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science, 16(6),* 440-444. DOI: 10.1111/j.0956-7976.2005.01554.x

Sangrigoli, S., & de Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry, 45(7),* 1219-1227. DOI: 10.1111/j.1469-7610.2004.00319.x

Sasson, N. J. (2006). The development of face processing in autism. *Journal of Autism and Developmental Disorders, 36(3),* 381-394. DOI: 10.1007/s10803-006-0076-3

Senju, A., Vernetti, A., Kikuchi, Y., Akechi, H. & Hasegawa, T. (2013). Cultural modulation of face and gaze scanning in young children. *PLos One, 8*(8): e74017. https://doi.org/10.1371/journal.pone.0074017

Spezio, M.L., Adolphs, R., Hurley, R.S.E., & Piven, J. (2007). Abnormal Use of Facial Information in High-Functioning Autism. *Journal of Autism and Developmental Disorders, 37,* 929-939. DOI: 10.1007/s10803-006-0232-9

Sue, D. W., & Sue, D. (1977). Barriers to effective cross-cultural counselling. *Journal of Counselling Psychology, 24(5),* 420-429.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research,* 168(3), 242- 249. https://doi.org/10.1016/j.psychres.2008.05.006 PMID: 19564050

Ueno, K., Utsuo, T., & Iinaga, K. (1991). *Picture vocabulary test*. Tokyo: Nihon Bunka Kagakusha. (In Japanese).

Walsh, J. H., Creighton, S. E., & Rutherford, M. D. (2016). Emotion Perception or Social Cognitive Complexity: What Drives Face Processing Deficits in Autism Spectrum Disorder? *Journal of Autism and Developmental Disorders, 46(2),* 615-623. doi: 10.1007/s10803-015-2606-3.

Webb, S. J., Neuhaus, E., & Faja, S. (2017). Face perception and learning in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *70*(5), 970-986.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2013). Face recognition deficits in autism spectrum disorders are both domain specific and process specific. *PLoS One, 8(9),* e74541. DOI: 10.1371/journal.pone.0074541

Wilson, C. E., Palermo, R., Burton, M. A., & Brock, J. (2011). Recognition of own- and other-race faces in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, 64(10), 1939-1954. doi: 10.1080/17470218.2011.603052.

Yi, L., Quinn, P. C., Fan, Y., Huang, D., Feng, C., Joseph, L., ... & Lee, K. (2016). Children with Autism Spectrum Disorder scan own-race faces differently from other-race faces. *Journal of Experimental Child Psychology, 141*, 177-186. doi: 10.1016/j.jecp.2015.09.011

Yi, L., Quinn, P. C., Feng, C., Li, J., Ding, H., & Lee, K. (2015). Do individuals with autism spectrum disorder process own-and-other-race faces differently*? Vision research, 107,* 124-132. DOI: 10.1016/j.visres.2014.11.021

Supporting Information

The information and analyses below relate to the larger sample of data collected as part of this study, before age-matched groups were made between the UK and Japan groups.

Participants

124 children participated in the study with 56 from the UK and 68 from Japan. Table S1 provides information on participant characteristics. All participants had normal or corrected-to-normal vision. In the UK sample, all children were recruited through local schools and advertisements. The age range for the UK sample was 6 to 13.4 years. The UK-ASD group consisted of 21 children (4 female, 17 male), with a mean age of 9.1 years (SD 2.19). Children with autism had all been previously diagnosed by experienced clinicians according to the DSM-IV criteria (American Psychiatric Association, 1994), and they all had a full statement of special educational needs and . or an education health and care plan (EHCP). The TD group consisted of 35 children (14 female, 21 male), with a mean age of 9.06 years (SD 1.11). Verbal ability was assessed using the British Picture Vocabulary Scale III (BPVS; Dunn & Dunn, 2009), and non-verbal ability was assessed with Ravens Colour Progressive Matrices (RCPM; Raven, Raven & Court, 1998b). The ASD group had a mean standardised BPVS score of 100.48 and a mean RCPM score 28.14. The TD group had a mean standardised BPVS score of 104.5, and a mean RCPM score of 28.57. There were no significant differences between ASD and TD groups for age, $t(54) = .071$, $p = .943$, $d = -.02$, standardised BPVS scores, $t(54) = -1.13$, $p = .265$, $d = -.3$, or RCPM score, $t(54) = -.30$, $p = .768$, $d = -.07$.

The age range for the Japanese sample was 6.6 to 13.6 years. In the Japanese sample, individuals with ASD were recruited through a medical university in the Tochigi area. The diagnosis of the participants with ASD was based on the DSM-IV) criteria (APA, 1994) and was confirmed by trained paediatric neurologists. Typically developing children were recruited from elementary, junior high, and high schools near in the Tochigi area. The ASD group consisted of 29 children (8 female, 21 male), with a mean age of 9.90 years (SD 1.52). The TD group was made up of 39 children (17 female, 22 male), with a mean age of 9.92 years (SD 2.03). Verbal ability was assessed using the Japanese version of the Peabody Vocabulary Test (Ueno, Utsuo & Iinaga, 1991), and non-verbal ability was assessed using the RCPM (Raven et al., 1998b). The ASD group had a mean Peabody scaled score of 10.83, and a mean RCPM of 30.72. The TD group had a mean Peabody score of 10.65, and a mean RCPM of 30.21. There were no significant differences between the TD and ASD groups for age, t(66) = .059, p = .953, d = .05, Peabody scores, $t(66)$ = .465, p = .643, d = .11, or Ravens, $t(66)$ = .509, p = .612, d = .12.

Comparison of the participants from both samples (Japan vs. UK) showed that the Japanese sample were significantly older than the UK sample (on average 9.91 years vs. 9.07), $t(122)$ = -2.5, p = .014, d = -.71, and they also had a significantly higher average RCPM score (30.43 vs. 28.43), $t(104.14)$ = 2.36, p = .020, d = .43. It was not possible to compare verbal ability given the different measures used. Given the differences between the samples[5], analyses of the face recognition data were carried out separately for the UK and Japanese groups, with comparisons made between cultural groups for the typicality of the ORE in ASD.

[5] Age and RCPM were found to correlate with accuracy in both samples, UK age, r(52) = .254, p = .03, RCPM, r(52) = .187, p = .09; Japan, Age, r(68) = .329, p = .003, RCPM, r(68) = .275, p = .01.

Results

Overall Accuracy for Face Recognition

UK sample: In terms of overall accuracy on the face recognition task, the UK-based ASD group mean was 70%, whilst the TD group mean was 73%. There was no significant difference between the ASD and TD groups for overall performance (see Figure S1), $t(54) = -1.02$, $p = .314$, $d = -.28$.

Japanese sample: The Japanese ASD group scored 80% on average, whilst the TD group scored 81% for overall mean accuracy. There was no significant difference between the ASD and TD groups in overall accuracy (see Figure S1), $t(66) = -.73$, $p = .469$, $d = -.177$.

Within each cultural sample there was no evidence of reduced accuracy for children with autism in terms of face recognition. Comparing cultural samples, the Japanese sample (81%) scored around 10% higher on average compared to the UK sample (72%) and this difference was significant, $t(122) = -5.50$, $p < .001$, $d = -.98$. This difference was not surprising given that the Japanese sample was older and had higher non-verbal ability, and these factors were found to be correlated with accuracy. Given the differences between the samples, analyses of the face recognition data were carried out separately for the UK and Japanese groups, with comparisons made between cultural groups for the typicality of the ORE in ASD.

Do children with autism show a reduced own-race advantage?

UK sample: Figure S2 represents the performance of the UK groups for accuracy across conditions. To examine performance across the task, a three-way mixed ANOVA was carried out using within subject factors face type (own and other) and difficulty (identity change, easy eyes, hard eyes, and hard mouth), and a between subjects factor of group (ASD and TD). There were main effects of face type; $F(1, 54) = 6.53$, $p = .013$, $\eta^2_p = .11$, difficulty; $F(2.56, 138.20) = 106.74$, $p < .001$, $\eta^2_p = .66$, and an interaction between face type and difficulty; $F(3, 162) = 5.77$, $p = .001$, $\eta^2_p = .10$. There was no main effect of group, $F(1, 54) = .912$, $p = .344$, $\eta^2_p = .017$, and no interactions of group and face type, $F(1, 54) = .018$, $p = .893$, $\eta^2_p < .001$, group and difficulty, $F(2.56, 138.20) = .438$, $p = .695$, $\eta^2_p = .008$, or group, face type, and difficulty, $F(3, 162) = .1.55$, $p = .208$, $\eta^2_p = .028$.

For the main effect of face type, the group means indicated that the UK sample was significantly more accurate at own race faces (M = 73.41%) than other race faces (M = 70.58%). For the main effect of difficulty, pairwise comparisons showed there was a significant difference in accuracy between all conditions (all *p's* < .001) apart from between the hard eye and hard mouth conditions ($p = 1.00$) (IC M = 92.02%, EE M = 79.88%, HE M = 58.72%, HM M = 57.35%). This indicated that the task became progressively more difficult between conditions, apart from between HE and HM which were equally difficult.

To examine the interaction of face type and difficulty paired samples t-tests were carried out. This revealed the interaction was driven by the identity change and easy eyes condition. In the identity change condition, the UK sample were significantly more accurate at own-race (95.09%) than other-race faces (88.84%), $t(55) = 2.99$, $p = .004$, d = -.38. They were also significantly more accurate at own-race (84.38%) than other-race (76.34%) in the easy eyes condition, $t(55) = 4.02$, $p < .001$, $d = -.48$ (Figure S2). There was no significant difference in the hard eyes condition between own-race (59.60%) compared to other-race faces (58.93%), $t(55) = -.234$, $p = .816$, $d = -.03$. This was also the case for the hard mouth condition (own-race: 55.80%; other-race: 59.60%), $t(55) = 1.53$, $p = .133$, $d = .23$.

These results show that although face type did impact performance (own-race easier than other-race particularly for the IC and EE conditions), this was not different between the ASD and TD groups. The ORA was further probed by analysing the differential scores, which not only allow for a more detailed look at the ORA but also an opportunity to explore heterogeneity. Differential scores can represent an ORA as they are made by subtracting total other-race performance for each condition from total own-race performance. Thus, a positive score indicates an ORA, whilst a negative score indicates an other-race advantage. Figure S3 shows the differential scores for the UK sample. One samples t-tests were carried out to compare the differential scores to a value of 0 (equal performance across face type conditions). In the UK ASD group, there was a significant ORA for the easy-eyes condition, $t(20) = 3.74$, $p = .001$. There was no significant ORA in the identity change condition, $t(20) = 1.56$, $p = .135$, hard eyes, $t(20) = .208$, $p = .837$, or hard mouth, $t(20) = -2.00$, $p = .059$, with the trend here indicating a potential other-race advantage. For the TD group, there was a significant ORA for the identity change condition; $t(34) = 2.60$, $p = .014$, and the easy-eyes condition; $t(34) = 2.17$, $p = .037$, but not the hard-eyes, $t(34) = .115$, $p = .909$, and hard-mouth, $t(34) = -.352$, $p = .727$.

Summary: By examining the interaction of face type and difficulty with the differential scores, the ORA for the UK sample seems to be present in the easier conditions (identity change and easy eyes) as opposed to the two more difficult conditions (hard eye and hard mouth). However, there seems to be more variable and heterogeneous performance in the

ASD group, with a trend towards an *other-race advantage* in the hard-mouth condition (indeed see error bars on Figure S3 representing heterogeneity).

Japanese Sample: Figure S4 illustrates the performance of the Japanese groups for accuracy across the conditions. There was a main effect of difficulty; $F(2.37, 156.65) = 147.96$, $p < .001$, $\eta^2_p = .69$, and an interaction between face race and difficulty; $F(2.01, 137.10) = 4.08$, $p = .018$, $\eta^2_p = .06$. However there was no main effect of face type, $F(1, 66) = 3.03$, $p = .086$, $\eta^2_p = .04$, or group, $F(1, 66) = .422$, $p = .518$, $\eta^2_p = .006$, no two-way interactions between group and difficulty, $F(2.37, 156.65) = .284$, $p = .790$, $\eta^2_p = .004$, face type and group, $F(1, 66) = 3.41$, $p = .069$, $\eta^2_p = .049$, and no three-way interaction between group, difficulty and face type, $F(2.08, 137.10) = .129$, $p = .886$, $\eta^2_p = .002$.

For the main effect of difficulty, pairwise comparisons showed that there was a significant decrease in accuracy between each difficulty level (all *p*'s < .001) (IC M = 97.58%, EE M = 90.80%, HE = 76.01%, HM = 60.25%). Paired samples t-test to unpick the interaction of face type and difficulty revealed this was driven by the hard eyes condition, where the Japanese sample was on average significantly more accurate at own-race (80.15%) than other-race faces (71.88%), $t(67) = 3.54$, $p = .001$, $d = .4$. There were no significant differences between own-race and other-race performance on the identity change condition (own-race = 97.79%, other-race = 97.61%), $t(67) = -.199$, $p = .843$, $d = -.03$; easy eyes condition (own-race = 90.07%, other-race = 91.73%), $t(67) = 1.10$, $p = .275$, $d = .11$; or hard mouth condition (own-race = 60.85%, other-race = 59.56%), $t(67) = -.431$, $p = .668$, $d = -.07$.

To further explore the ORA, an analysis of differential scores was conducted. Figure S5 shows the data for differential scores for the Japanese sample. One sample t-tests revealed that for the ASD group there was a significant ORA for the hard-eyes condition; $t(28) = 2.19$, $p = .037$. However there no significant ORA for the identity change, $t(28) = -1.00$, $p = .326$; easy eyes, $t(28) = -1.61$, $p = .118$, and hard mouth conditions, $t(28) = -.434$, $p = .668$. The TD group likewise showed a significant ORA on the hard eyes condition, $t(38) = 2.79$, $p = .008$, but no significant ORA for the identity change, $t(38) = 1.00$, $p = .324$; easy eyes, $t(38) = -.154$, $p = .878$; and hard mouth conditions, $t(38) = 1.05$, $p = .302$.[6]

[6] Neither Age, RCPM nor verbal ability correlated with differential percent scores for any group within either cultural sample.

		ASD	TD
Gender (M/F)	**UK**	17/4	21/14
	Japan	21/8	22/17
Age (Years) M (SD)	**UK**	9.10 (2.19)	9.06 (1.77)
	Japan	9.90 (1.52)	9.92 (2.03)
Peabody M (SD)	**UK**	N/A	N/A
	Japan	10.83 (1.38)	10.65 (1.64)
Ravens M (SD)	**UK**	28.14 (6.13)	28.57 (4.63)
	Japan	30.72 (3.73)	30.21 (4.44)
BPVS M (SD)	**UK**	100.48 (13.34)	104.51 (12.78)
	Japan	N/A	N/A

Table S1: A comparison of the descriptive statistics of the ASD and TD groups of the UK and Japanese samples

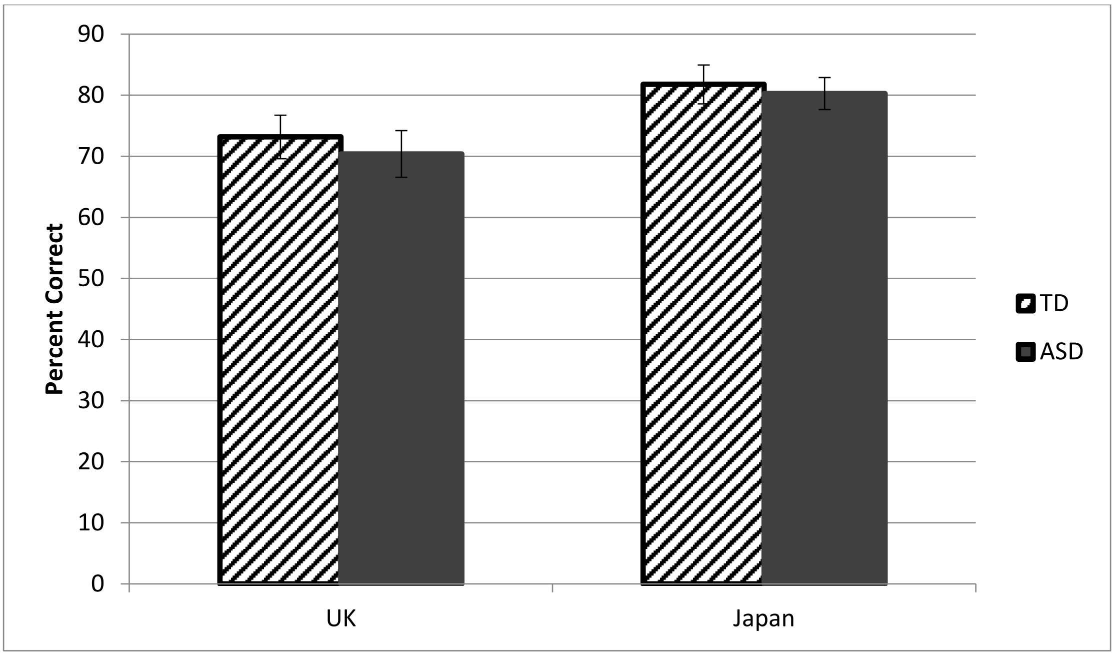


Figure S1: Overall accuracy for the ASD and TD groups of both samples. Error bars represent +/- 2 SE of the mean.

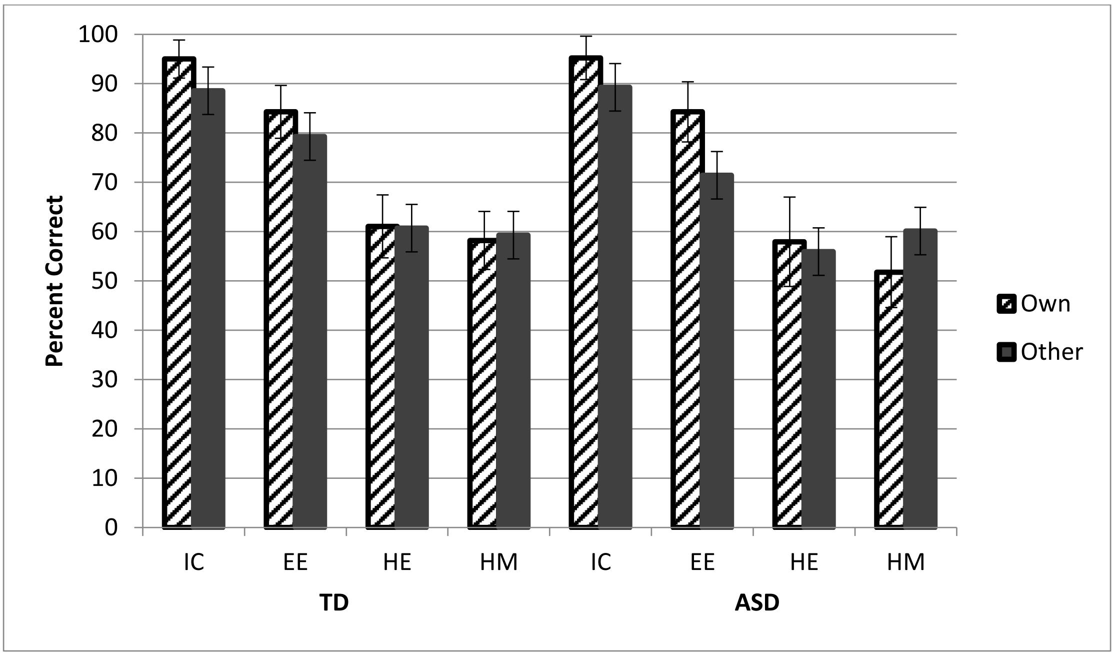


Figure S2: UK own and other-race performance for each difficulty condition for the ASD and TD groups. Error bars represent +/- two standard errors.

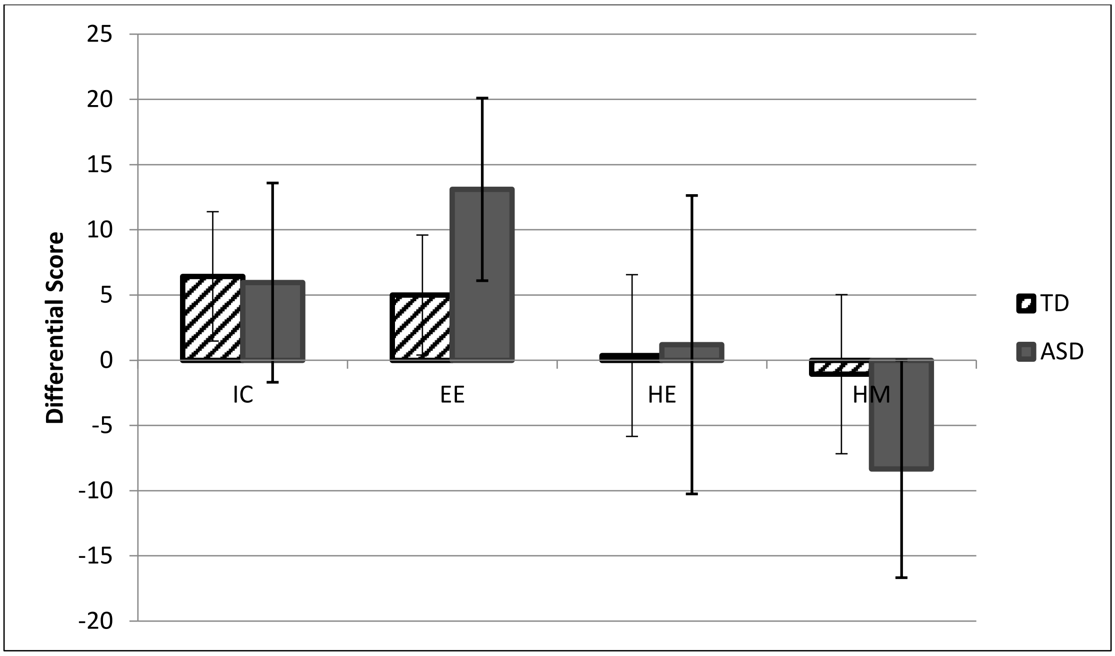


Figure S3: UK differential percent scores for the ASD and TD groups. Error bars represent +/- two standard errors.

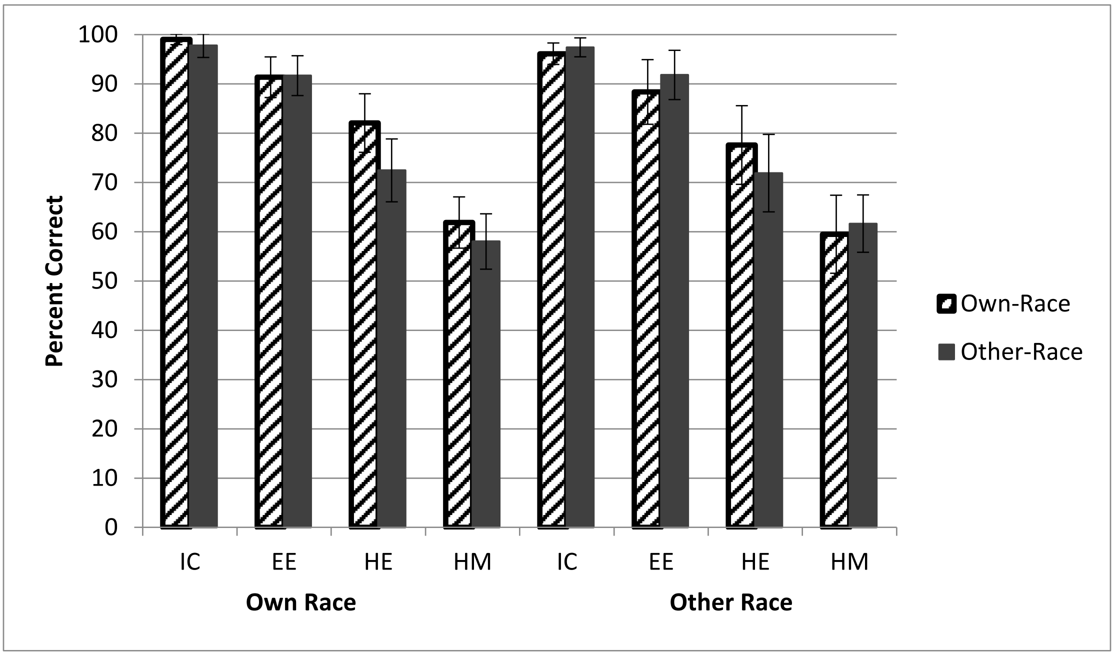


Figure S4: Japan own and other-race performance for each difficulty condition for the ASD and TD groups. Error bars represent +/- two standard errors.

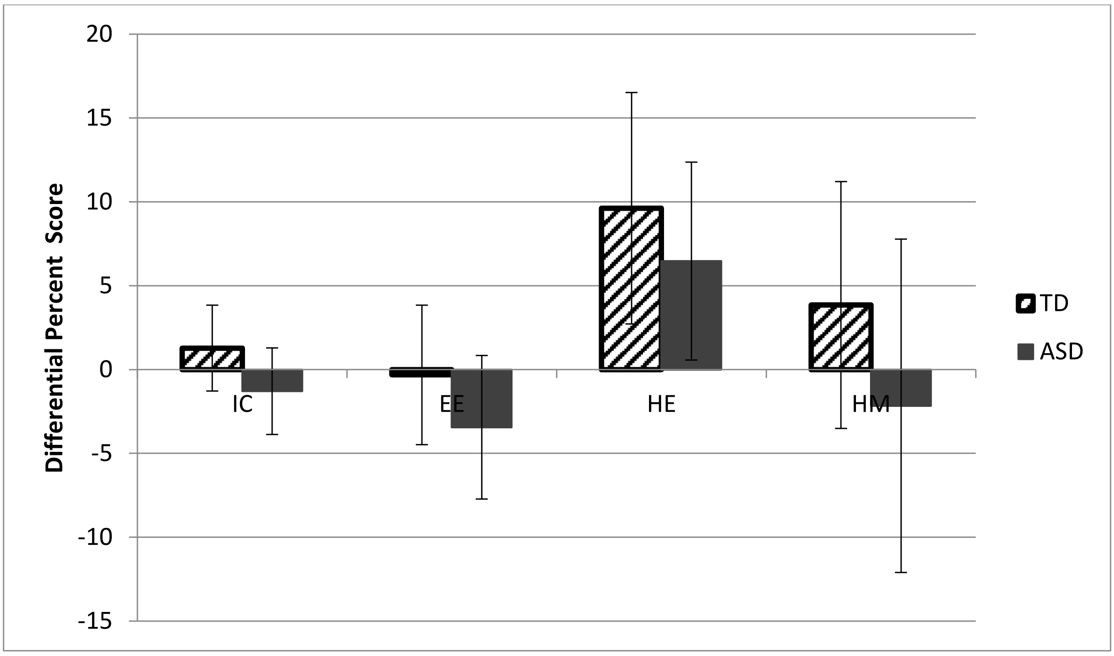


Figure S5: Japanese sample differential percent scores for the ASD and TD groups. Error bars represent +/- two standard errors.