Journal of Child Psychology and Psychiatry 45:7 (2004), pp 1258–1274 doi: 10.1111/j.1469-7610.2004.00322.x

Exploring the Williams syndrome face-processing debate: the importance of building developmental trajectories

Annette Karmiloff-Smith,[1] Michael Thomas,[2] Dagmara Annaz,[2] Kate Humphreys,[1] Sandra Ewing,[1] Nicola Brace,[3] Mike Van Duuren[4], Graham Pike,[3] Sarah Grice,[1] and Ruth Campbell[5]

[1]Neurocognitive Development Unit, Institute of Child Health, London, UK; [2]School of Psychology, Birkbeck College, London, UK; [3]The Open University, UK; [4]King Alfred's College, Winchester, UK; [5]Department of Human Communication, University College London, UK

Background: Face processing in Williams syndrome (WS) has been a topic of heated debate over the past decade. Initial claims about a normally developing ('intact') face-processing module were challenged by data suggesting that individuals with WS used a different balance of cognitive processes from controls, even when their behavioural scores fell within the normal range. Measurement of evoked brain potentials also point to atypical processes. However, two recent studies have claimed that people with WS process faces exactly like normal controls. **Method:** In this paper, we examine the details of this continuing debate on the basis of three new face-processing experiments. In particular, for two of our experiments we built task-specific full developmental trajectories from childhood to adolescence/adulthood and plotted the WS data on these trajectories. **Results:** The first experiment used photos of real faces. While it revealed broadly equivalent accuracy across groups, the WS participants were worse at configural processing when faces were upright and less sensitive than controls to face inversion. In Experiment 2, measuring face processing in a storybook context, the face inversion effect emerged clearly in controls but only weakly in the WS developmental trajectory. Unlike the controls, the Benton Face Recognition Test and the Pattern Construction results were not correlated in WS, highlighting the different developmental patterns in the two groups. Again in contrast to the controls, Experiment 3 with schematic faces and non-face stimuli revealed a configural-processing deficit in WS both with respect to their chronological age (CA) and to their level of performance on the Benton. **Conclusion:** These findings point to both delay and deviance in WS face processing and illustrate how vital it is to build developmental trajectories for each specific task. **Keywords:** Williams syndrome, configural, holistic, featural, face processing, developmental trajectories, progressive modularisation of function.

In this article, we investigate the emergence of face recognition abilities in a rare genetic developmental disorder, Williams syndrome. The capacity for species recognition is one of the most fundamental abilities across the animal kingdom. One might therefore expect this to be a strong candidate for an innate ability in the case of the human infant. Yet the past couple of decades have revealed, at both the behavioural and brain levels, that the recognition of faces is a very gradual developmental process in both humans and other species such as the chick (Johnson & Morton, 1991). Indeed, brain localisation and specialisation in the processing of human faces, i.e., the gradual modularisation of function over developmental time (Karmiloff-Smith, 1992, 1998), extends very progressively across the first 12 months of life and beyond (Johnson & de Haan, 2001). If infants start with anything resembling an innately specified template, it is unlikely to be face specific but rather in the form of a T-shape in which more information at the top of a stimulus parsed as an object is particularly attractive to the young infant's visual system (Simion, Macchi Cassia, Turati, & Valenza, 2001). With developmental time, however, the human face itself becomes increasingly the preferred stimulus (Johnson & Morton, 1991), enhanced not only by the massive input of faces but also by the fact that faces form a crucial site of attention for the social interactional patterns that develop over the first months of life.

Imaging experiments have shown that young infants' brains initially process upright human faces, inverted human faces, monkey faces and objects all in a relatively similar way across both hemispheres (Johnson & de Haan, 2001; de Haan, 2001). However, with development, brain processing of human upright faces becomes increasingly specialised and localised to the fusiform gyrus in the right hemisphere (Passarotti et al., 2003). Despite these developmental data pointing to very progressive specialisation and localisation of face processing, some theorists claim that the human brain is prespecified with an independently-functioning face-processing module. Such claims are based on the fact that *adult* patients can present with prosopagnosia, i.e., a selective inability to recognise familiar faces, despite showing no obvious impairments elsewhere (Bruyer et al., 1983; de Renzi, 1986;

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Farah, Levinston, & Klein, 1995; Temple, 1997). So, there is still a debate concerning the extent to which face recognition abilities are part of the hardwired functional architecture of the infant brain versus the extent to which these emerge in adults mainly as a product of development. Can developmental disorders address this question, particularly those where face recognition appears to exceed general cognitive ability, such as in Williams syndrome?

Williams syndrome

Williams syndrome (WS) is a genetic disorder in which some 25 genes are deleted on one copy of chromosome 7, leading to serious deficits in spatial cognition, number, planning and problem solving (see Donnai & Karmiloff-Smith, 2000, for full details). IQ scores are in the 50s to 60s range (Udwin & Yule, 1991). Of particular interest to cognitive neuroscientists is the fact that two domains – language and face processing – show particular behavioural proficiency compared to the general levels of intelligence reached by this clinical group. Indeed, WS scores on some language and face-processing tasks fall in the normal range. When such findings first arose, they were heralded as demonstrating that, in the case of face processing, for instance, WS presents with an 'intact' or 'preserved' face-processing module (e.g., Bellugi, Sabo, & Vaid, 1988; Bellugi, Wang, & Jernigan, 1994; Wang, Doherty, Rourke, & Bellugi, 1995; though see Bellugi, Lichtenberger, Mills, Galaburda, & Korenberg, 1999, for more recent discussion). While some have rejected these claims (Karmiloff-Smith, 1997, 1998; Deruelle, Mancini, Livet, Cassé-Perrot, & de Schonen, 1999), others maintain that people with the syndrome display normal face processing (e.g., Deruelle, Rondan, Mancini, & Livet, 2003; Tager-Flusberg, Plesa-Skwerer, Faja, & Joseph, 2003). This is of course tantamount to claiming that face processing *develops* normally in WS.

One of the problems with many of the WS face-processing studies is that terms like featural/piecemeal/componential/local/analytical versus configural/holistic/global/gestalt have been used interchangeably, as if they were synonymous, and thus have not been adequately specified. In this paper, we use the term 'featural' to refer to the ability to identify faces based on individual features (eyes, nose, mouth, chin), and the term 'configural' to refer to the ability to differentiate faces based on sensitivity to the spatial distances amongst internal features, i.e., second-order relational information. Configural face processing is associated with maturity of face recognition encoding and 'expert' recognition in adults. By contrast, the term 'holistic' is deemed to cover the gluing together of facial features (and hairline) into a gestalt, without necessarily conserving the spatial distances between features (Maurer, Le Grand, & Mondloch, 2002; Tanaka & Farah, 1993). In other words, the capacity to process information holistically does not involve the processing of second-order relational information. We start with a review of previous work on face processing in Williams syndrome, followed by three new experiments with this clinical population.

Previous WS face-processing studies

There are no experimental studies of face processing in infants with WS to complement the studies of healthy infants discussed in our introduction. However, some observational work, as well as experiments indirectly tapping face processing, revealed that infants with WS spend significantly more time focused on faces than on objects (Bellugi, Lichtenberger, Jones, Lai, & St. George, 2000; Laing et al., 2002; Mervis & Bertrand, 1997). This has led many authors to assume that the WS infant's inordinate attention to faces explains why adults with WS end up achieving good behavioural scores on some face-processing tasks. But does such early attention to faces necessarily lead to configural processing in older individuals with WS?

The initial claims that adolescents and adults with WS exhibit face recognition skills that are 'intact'/'spared' (i.e., developed normally) were based on findings that performance on the standardised face-processing tasks like the Benton Facial Recognition Test and the Rivermead Face Memory Task was at normal or near normal levels. However, several studies subsequently challenged the notion of an 'intact' face-processing module and suggested that people with WS achieve their normal scores by resorting to different strategies from controls. Karmiloff-Smith (1997) reported that on a face-matching task, adult participants with WS ($N = 10$) did not differ from chronological-age-matched controls with respect to featural analysis, but were significantly worse when items necessarily required configural analysis, i.e., taking account of second-order relations. These preliminary findings, the result of acknowledged post-hoc analyses, gained support from a later study by Deruelle et al. (1999). Twelve children and adults with WS, aged between 7 and 23, were compared against chronological-age (CA) and mental-aged (MA) matched controls in a task requiring participants to decide whether two pictures of faces were the same or different when presented in upright and inverted conditions (their second experiment). The clinical group was less subject to an inversion effect than the controls. The authors explained these results by a greater reliance of the WS participants on featural analysis in both the upright and inverted conditions, whereas the controls used predominantly featural processing for the inverted faces and configural processing for the upright faces. This led the authors to speculate that WS face processing is not merely delayed but follows a different developmental pathway. In a third experiment, Deruelle et al. (1999) investigated the processing of configurally and featurally modified schematic faces and geometric shapes. Yet again, the CA and MA matches produced significantly fewer errors than the WS group on configural items, but no differences emerged with respect to the featural ones. The 1999 Deruelle et al. study adds further support to the claim that people with WS are biased to process featural over configural information, regardless of the type of facial (real faces or schematic faces) or geometric stimuli.

Two recent papers have, however, challenged this now rather general conclusion and claimed that people with WS process faces in exactly the same way as controls (Deruelle et al., 2003; Tager-Flusberg et al., 2003).

Deruelle et al. (2003) sought to explore holistic face processing in WS. They compared 12 children and adolescents with WS (aged 6 to 17 years) against controls in their ability to match faces to either a low or high spatial frequency filtered target face. Two control groups were used, either matched individually on chronological age or on overall mental age. All groups tended to find face matching easier to a low spatially filtered target face (i.e., a face in which the broad patterns of light and dark were preserved but fine detail was lost) than a high spatially filtered face. WS performance fell between the CA and MA control groups but was not significantly different from either. While Deruelle et al. took these results to indicate that face-processing abilities develop normally in children with WS, the sensitivity of responses to the spatial frequency manipulation was unchanging with age across all the groups, despite the wide age range. This implies that, by the age of testing, any developmental change had already plateaued in terms of holistic face processing, making it impossible to assess the implications for the WS participants on the developmental trajectory.

Tager-Flusberg et al. (2003) have also recently argued in favour of normally developing face processing in WS. That study had the merit of using a considerably larger sample than previous work, including 47 adolescents and adults with WS (aged 12 to 36 years) and 39 CA-matched controls. They were tested on the standardised Benton Face Recognition Test used in previous studies, as well as on a part–whole paradigm. This latter task involved matching individual face parts in two conditions, either in isolation or in the context of a whole face. The authors predicted that if the WS group were less influenced by the overall context of a face, they should show a reduced difference between the two conditions, whereas the controls should be aided in the recognition of individual features by their presentation in the context of a face. Face orientation was also varied to assess whether the ability to use the whole-face context was disrupted by inversion to the same extent in the control and WS groups. While the overall accuracy was better in the CA-matched control group than in the WS group, the presence or absence of surrounding face context had the same effect for upright presentation on controls and the clinical group. Tager-Flusberg et al. present their results as challenging previous WS data that pointed to atypical face processing and argue that earlier studies were underpowered because of small *N*s. However, it is not clear that their task addresses the same aspect of face processing as previous studies. Their part–whole task (taken from Tanaka & Farah, 1993) concerns the processing of individual features recognised in isolation or in the context of a whole face. This taps, as the authors recognise, the processing of the face gestalt, i.e., first-order holistic processing, rather than second-order configural processing. Indeed, throughout their article, Tager-Flusberg et al. contrast featural processing with 'holistic' processing rather than with configural processing. Yet the Deruelle et al. (2003) study indicates that individuals with WS are not markedly impaired on holistic processing and none of the existing research, including our own, has claimed that individuals with WS are incapable of first-order holistic processing. The debate is about second-order configural processing. Of importance, too, is the fact that in the Deruelle et al. study holistic processing assessed by spatial frequency manipulations exhibited no developmental change across the age range employed by Tager-Flusberg and colleagues. Similarly, when these researchers examined developmental change across their (wide) age range, they also found no correlation between age and performance in either the WS or CA groups. If the part–whole task targets holistic processing, and if holistic processing is at ceiling in both normal and clinical populations in the age ranges examined, then the conclusions than can be drawn regarding the typicality or atypicality of face-processing development in WS are obviously limited.

As noted, claims of atypicality in WS face processing revolve around the extent to which these individuals make use of *second-order configural relations* when recognising faces. This is a capacity that emerges over developmental time in healthy controls. Therefore, it is the inversion condition used by Tager-Flusberg et al. that is potentially more informative here, since inversion causes disruption particularly to configural processing (Diamond & Carey, 1986). Unfortunately, Tager-Flusberg and colleagues' WS data are uninterpretable in this regard, because the clinical group is at floor on inverted stimuli. By contrast, two previous studies from separate laboratories have found significantly less difference between upright and inverted face recognition in WS groups compared to controls (Deruelle et al., 1999; Rossen, Jones, Wang, & Klima, 1995), again supporting the claim that this clinical population is atypical with regard to configural processing. The possibility that inconsistent findings in the WS face-processing literature are a result of confounding holistic and configural processing is one to which we will return. For the moment, we focus on further investigating configural processing in WS face recognition.

The importance of developmental trajectories. Since the major theoretical dispute concerns whether face recognition develops normally in WS, it will be our contention that explanations must be couched in terms of developmental trajectories (see discussions in Karmiloff-Smith, 1998, and Thomas & Karmiloff-Smith, 2002). Indeed, this stance will influence our analytic techniques. It will cause us to move from comparisons of the WS group against individually-matched controls which we and other groups have used in the past (where the relation of performance to age is discarded in analyses once control participants are selected) to the construction of functions specifically linking performance with age. Our comparisons will relate performance to chronological age in both the control and clinical samples and, in the latter, also between our face-processing tasks and developmental stages on a number of standardised tests, including face recognition, visuospatial processing, and language (see Thomas et al., 2001, for a similar analytical approach). We believe that our new trajectory approach to disorders offers more insight with respect to the way in which development may have proceeded *over time* in a deviant fashion, even though the behavioural proficiency, measured by matching controls at a specific CA or MA, may end up similar to controls. Moreover, individual variability is as much a problem for individual matching as it is for building trajectories. Each participant exhibits individual variation from each group norm. In addition, we try to show in this paper that comparing trajectories has a marked advantage over individual matching with regard to MA comparisons. In an individually matched comparison, everything rides on the choice of MA measure and this is typically theoretically laden. However, if one builds a task-specific typical developmental trajectory, one can then evaluate whether the atypical group fits on the typical developmental trajectory according to a range of metrics. These can be theoretically neutral (does each individual fit *anywhere* on the trajectory), or according to CA, or according to a variety of MA measures. Finally, where wide age ranges are a problem for studies matching on CA or MA and age is ignored in subsequent analyses, the wide age range can actually be exploited positively when building developmental trajectories.

Our aim in this article is threefold. In our first experiment we aim to produce clearer evidence of the configural processing deficit and reduced inversion effect in a WS group, by using real faces in a task specifically tailored to this purpose. In our second experiment, we concentrate on the inversion effect. We establish a cross-sectional, task-specific developmental trajectory for a wide age range of normal controls on a task that embeds the recognition of inverted faces in the naturalistic context of a storybook. We then derive and contrast the trajectory on this task for a group of individuals with WS. In Experiment 3, we concentrate on sensitivity to configural vs. featural transformations of schematic faces and geometric patterns, once more building and contrasting developmental trajectories for the typical and atypical groups. In contrast to the findings of Deruelle et al. (2003) and Tager-Flusberg et al. (2003), our tasks allow us to chart developmental change in performance across our age ranges in both groups. In each case, we will reveal patterns of atypicality in WS face recognition over developmental time, even when accuracy levels are broadly similar.

Hypotheses. Experiment 1: If face recognition in WS has developed normally, individuals with WS should show no difference in accuracy compared to controls in discriminating a target face from a featurally or configurally transformed version of the target; moreover, if WS face processing has developed normally, then performance should be similarly influenced by the nature of the transformation and by inversion.

Experiment 2: When required to recognise contextualised faces in a storybook task, presented in upright or inverted orientation, we expect controls to become increasingly accurate and rapid with age with respect to upright faces, but also to show increased sensitivity to inversion (since this disrupts configural processing). If face recognition in WS has developed normally, the clinical group should show the same pattern as the controls with increasing chronological age. By contrast, if the developmental trajectory is simply delayed, then performance should be predicted by developmental age on the standardised Benton face recognition test, or on a standardised test of visuospatial ability (but possibly not on a standardised test of vocabulary). Finally, if face processing in WS follows a deviant trajectory, then these predictions should not hold.

Experiment 3: We expect controls to demonstrate an emerging skill in detecting configural transformations in upright faces with increasing age. If face recognition in WS has developed normally, the clinical group should show the same pattern as the controls when matching schematic face and non-face stimuli transformed configurally or featurally. By contrast, if the developmental trajectory is merely delayed, then it should be predicted by developmental age on the Benton task, or on a standardised test of visuospatial ability. Finally, if face processing in WS follows a deviant trajectory, then these predictions should not hold.

Experiment 1: Investigating configural and featural processing of real faces

We begin by assessing the presence of differences in configural processing and in the inversion effect in

WS using real faces in a test specifically designed for this purpose by Mondloch and colleagues (Mondloch, Le Grand, & Maurer, 2002).

Method

Participants.[1] Fourteen adult participants with WS were tested on this task. Two did not fully understand the instructions and their data were excluded from the analyses. The 12 remaining participants had a mean CA of 30;0 (SD: 11;11, range: 16;3–51;0). Language ability was assessed using the Peabody Picture Vocabulary test (Dunn & Dunn, 1981) for which their mean test age was 13;10 (SD: 6;7, range: 8;2–30;8). Visuospatial ability was assessed using the Pattern Construction subtest of the British Abilities Scale (BAS-II; Elliot, Smith, & McCulloch, 1996). The mean spatial test age of the WS group was 5;4 (SD: 1;7 range 3;7–8;9). Thus the WS group exhibited the characteristic disparity between these two abilities (language greater than visuospatial skill: paired t-test, $t = 5.42$, $df = 11$, $p < .001$), alongside overall delay. A control group was individually matched to each participant in the WS group, based on chronological age, gender and socio-economic status. The choice of a chronological age match was based on the claims that face processing is an independently-functioning, intact ability in this clinical population, as well as to equate to the best extent possible for life experience with faces in the two groups. MA controls could have been interesting, too, but were not our main focus here. The CA matches were on average within 4.5 months of each of the participants with WS, with the mean CA of the control group being 29;11 (SD: 11;6, range 16;6–51;0). A comparison of the CAs of the control and WS groups revealed no significant difference (paired t-test, $t = .19$, $df = 11$, $p = .856$).

Stimuli. A photo of a real face (called Jane) was used to create featural and configural sets of new faces. In the featural set, new faces were created by replacing the original features (eyes and mouth) with the features of different faces. In the configural set, features were moved up or down within the face contour, or moved closer together or further apart in relation to the original positions of the features (for more detailed description of the task used with a different population, see Mondloch, Le Grand, & Maurer, 2002; note the 'contour' condition of that study was not employed here).

Procedure. The procedure employed a well-tested paradigm for directly differentiating between featural and configuring processing of real faces (Mondloch et al., 2002). Participants were presented sequentially with two faces and asked to determine whether the two faces were the same or different. Trials were blocked into upright and inverted trials, and within those, separate blocks where faces were featurally altered or configurally altered (referred to as 'spacing' in Mondloch et al.). Trials were blocked to encourage participants to adopt specific face-processing strategies (Mondloch et al., 2002). The testing session began with 12 practice trials, to ensure that all participants understood the instructions and meaning of the words 'same' and 'different'. During the task proper, each participant was presented on a computer laptop with 30 trials from the featural and configural sets respectively. For each participant, the upright block was always presented before the inverted block. The order of configural and featural blocks within these was counterbalanced. Each block consisted of 15 'same' and 15 'different' randomised trials. During each trial, the target face was presented for 400 ms and the second face, to which the participant had to respond on the keyboard with 'same' or 'different', was displayed until the response button was pressed. The inter-stimulus interval was 300 ms, a delay chosen to prevent apparent-motion cues from the presentation of a different face (Mondloch et al., 2002). Bright yellow Velcro pads were stuck on the two relevant keys to make it easy for participants to respond. For each trial a target face was followed by a test face that could either be identical to the target face or transformed configurally or featurally. The dependent variables were reaction time and accuracy.

For the analysis, we divided the task into two components: *Identity Recognition* (for all items where no change had been made between target and test face and the participant correctly responds 'same'), and *Difference Detection* (where the participant correctly spots that the test face differs from the target, either due to a configural or featural transformation, and responds 'different'). We consider the two sets of responses separately, because transformations are only relevant to difference detection.

Results

Identity Recognition. A comparison of accuracy levels in identity recognition revealed no significant difference between the groups (ANOVA: main effect of group: $F(1,22) = 1.38$, $p = .253$, $\eta_p^2 = .059$). In addition, both groups performed equally well on upright and inverted faces. Prior to analysis, response time data were cropped so that outliers that fell two standard deviations away from each individual's mean time were eliminated. This removed 4.7% of the data points from the WS group and 4.4% of the data points from the CA control group. For identity recognition, individuals with WS tended to respond more slowly than CA controls, but this effect was not reliable ($F(1,22) = 3.38$, $p = .081$, $\eta_p^2 = .132$). Where the WS were slower (in mean scores), they were also less accurate than the control group, suggesting that there were no speed–accuracy trade-offs at work.

Whether identity recognition trials fell within configurally transformed or featurally transformed blocks had

[1] All participants with WS across the three experiments (a different opportunity sample each time because of the great distances) had been diagnosed clinically and by means of the FISH probe for the deletion of the elastin gene. Because of the rarity of Williams syndrome and the size of the United Kingdom, numbers are necessarily somewhat low, ranging from 12 to 17 participants with WS per experiment. However, this is well above many of the published papers on this syndrome and is comparable, for instance, to the study of Deruelle et al. (2003), which was viewed by these authors as sufficient to make claims about 'normal' development in WS face processing.

Table 1 Means and standard errors (SE) for accuracy (%) and response times (ms) in Identity Recognition and Difference Detection

Identity Recognition Group	Trial block: 'Featural' Upright Mean	(SE)	'Featural' Inverted Mean	(SE)	'Configural' Upright Mean	(SE)	'Configural' Inverted Mean	(SE)
WS								
Accuracy	72%	(6%)	67%	(10%)	81%	(4%)	74%	(9%)
RT	1702	(355)	1531	(249)	1630	(377)	1565	(306)
Control								
Accuracy	80%	(6%)	83%	(4%)	74%	(5%)	87%	(5%)
RT	880	(101)	1011	(136)	1115	(147)	992	(87)

Difference Detection Group	Transformation: Featural Upright Mean	(SE)	Featural Inverted Mean	(SE)	Configural Upright Mean	(SE)	Configural Inverted Mean	(SE)
WS								
Accuracy	86%	(5%)	74%	(7%)	51%	(6%)	31%	(8%)
RT	1701	366	1569	333	1766	371	1895	444
Control								
Accuracy	85%	(5%)	83%	(6%)	75%	(4%)	29%	(7%)
RT	873	73	874	62	1033	99	1136	112

no effect on performance (effect of transformation block: errors, $F(1,22) = 1.02$, $p = .323$, $\eta_p^2 = .044$; response times, $F(1,22) = .43$, $p = .519$, $\eta_p^2 = .019$). No significant interactions of transformation block emerged with any of the other variables, implying that trial blocking of featurally vs. configurally transformed faces did not induce specific face-processing strategies sufficient to affect identity recognition. The mean accuracy levels and response times are provided in Table 1.

Difference Detection. The difference detection condition includes the additional variable of manipulation type: configural or featural. Accuracy and response times for difference detection are also included in Table 1. Once more, there was no significant difference in accuracy levels between the groups (effect of group: $F(1,22) = 2.03$, $p = .168$, $\eta_p^2 = .085$). Analysed in isolation, the CA control group exhibited a characteristic pattern in the difference detection task, whereby it turned out to be harder to detect configurally changed faces than featurally changed faces (main effect of transformation: $F(1,11) = 26.89$, $p < .001$, $\eta_p^2 = .710$), and whereby inverting the face added to the difficulty only for configurally changed faces (interaction of transformation × orientation: $F(1,11) = 61.97$, $p < .001$, $\eta_p^2 = .849$). Overall, the WS group experienced the same effects of transformation, i.e., configurally transformed faces were not additionally harder for the participants with WS (main effect of transformation in the WS group: $F(1,11) = 33.61$, $p < .001$, $\eta_p^2 = .753$, between-group comparison, interaction of group × transformation: $F(1,22) = .57$, $p = .457$, $\eta_p^2 = .025$). However, the clinical population did exhibit a differential effect of inversion on the face transformation compared to the control group. In particular, they demonstrated a larger inversion effect on featurally transformed faces than the CA control group and a smaller inversion effect on configurally transformed faces (between-group comparison, interaction of group × transformation × orientation: $F(1,22) = 14.73$, $p = .001$, $\eta_p^2 = .401$). Inspection of Table 1 (lower panel) suggests that this interaction was driven most strongly by a greater disparity between WS and CA control groups in detecting configurally transformed upright faces. Indeed, in a direct comparison of detecting configural transformations in upright faces, the WS group were significantly worse than controls, but showed no difference when these stimuli were inverted (independent-sample *t*-tests comparing groups, Upright: $t(22) = 3.34$, $p = .003$; Inverted: $t(22) = .16$, $p = .878$). In other words, the normal configural expertise for upright faces found in the CA-matched controls was not apparent in the WS group, despite broadly equivalent overall levels of accuracy. The WS group responded more slowly than the CA control group but, as with Identity Recognition, this did not reach significance (main effect of group: $F(1,22) = 4.09$, $p = .056$, $\eta_p^2 = .157$).

Discussion

The clinical group turned out to be as accurate as the CA controls on both identity recognition and difference detection and, although they tended to respond more slowly, the difference was not significant in either condition. By these measures alone, one might conclude that face recognition had developed normally in the WS group. However, the groups diverged in other important ways, in particular on transformations in the difference detection task. While configural transformations were harder to detect than featural transformations, the key condition that separated the groups was performance on upright configural faces, where controls were significantly more accurate. Sensitivity to second-order configural differences in upright faces is a hallmark of face recognition expertise (Yin, 1969). It is most disrupted by inversion, and the WS group demonstrated a qualitatively different response to inversion that stemmed particularly from weaker performance on upright configurally transformed faces.

Experiment 2: Face recognition in a story-supported task

In Experiment 2, we consider the progressive developmental emergence of the inversion effect, using a task that embeds the recognition of inverted faces in the more naturalistic setting of a storybook. Here, we introduce a novel approach to controls, by building a full task-specific developmental trajectory.

Method

Participants. Seventeen adolescent and adult participants with WS were tested. Three participants failed to correctly identify either a single upright face or a single inverted face and were excluded from subsequent analyses. The 14 remaining participants with WS had a mean CA of 26;3 (SD: 11;11, range: 12;0–54;10). Language ability was assessed using the British Picture Vocabulary Scale (BPVS) (Dunn, Whetton, & Pintilie, 1997), for which the mean test age was 11;2 (SD: 3;6, range: 7;4–17;6). Visuospatial ability was assessed using the Pattern Construction subtest of the British Abilities Scale (Elliot et al., 1996). The mean test age of the WS group was 5;6 (SD: 1;2, range 3;4–8;9). Thus, this new WS group exhibited the characteristic disparity between these two abilities (language greater than visuospatial skill: paired *t*-test, $t = 7.51$, $df = 13$, $p < .001$), alongside overall delay. Participants were also given the Benton Face Recognition Test. Mean Benton score was 42.2 (SD: 4.5, range: 35–51), corresponding to performance within the normal adult range (41–54; Benton, Hamsher, Varney, & Spreen, 1983). Using the age norms from the Benton, these scores were translated into age equivalents (Benton et al., 1983), taking a ceiling score to be reached at 14 years. This gave the WS group a mean Benton age equivalent of 11;4 (SD: 2;7, range: 6;6–14;0).

Data from 111 control children were also analysed. These data were collected by Brace et al. (2001), with group means derived for separate age bins. We reanalysed the data for this sample, using each child's CA to build a task-specific cross-sectional developmental trajectory. The control group had a mean CA of 8;0 (SD: 2;7, range: 2;8–11;5). There were 12 children between 2;8 and 4;4, 20 children between 5;2 and 6;11, 26 children between 7;2 and 8;8, 28 children between 9;2 and 10;2, and 25 children between 11;1 and 11;5. This age range enabled us to build a full trajectory of typical developmental changes on this specific task.

Stimuli. The stimuli were taken from Brace et al. (2001) and were modified for use with a touch-screen monitor using Superlab 2.0. The stimuli consisted of two parts: a Storybook and a computer game. The Storybook was a hand-painted story about two boys, called Tom and Jamie. One of the boys is kidnapped by a witch and taken to her castle. The witch turns the boy into a variety of objects, such as a robot that retains only the boy's face, and hides him in amongst 8 other boys that she has kidnapped. The only way for the other boy to rescue his friend is to play a game of hide and seek in order to spot his friend amongst the other boys/objects, which are either upright or hung upside-down. In the first two pages of the book, pictures of Jamie and Tom are present, whereas in the next five pages the story continues without any pictures of the boys, to ensure that subsequent recognition of the faces is delayed by about three minutes. The hide and seek computer game includes upright and inverted pictures of one of the two target faces (Tom or Jamie) among 8 distracter faces. Two versions of the task were run with different target faces. Each participant saw one of the two versions (for further details, see Brace et al., 2001).

Procedure. In the first part of the study the experimenter (or the participant if s/he read easily) read a story aloud, during which the participant was asked to point to the pictures of the two boys and to repeat their names after the experimenter. On completion of the first part of the story, the participant was asked if s/he would like to play a computer game of finding the lost boy (Tom or Jamie). Eight trials were run, including two practice trials. For each trial, a picture was presented on the touch-screen with the target face hidden amongst 8 distractor faces of varying similarity to the target face. The position and orientation of the target face within the array of 9 faces was systematically varied. Once the detection game was completed, the story reading was continued to achieve a happy ending. Participants' data were only included if they correctly recognised at least one upright and one inverted instance of the target face. In most cases, performance accuracy was much better.

Results

The wide age and ability range of the WS and control groups was exploited to generate developmental trajectories on the story task, i.e., to create a function relating increase in performance (either accuracy or response time) against increasing age. For the WS group, 'age' was for each analysis either their chronological age, their Pattern Construction equivalent, their Benton equivalent or their BPVS equivalent. The WS developmental trajectory was compared against the developmental trajectory for the controls, constructed from over one hundred typically developing children between 2;8 years and 11;5 years.

WS performance on Benton Facial Recognition Test. First, we establish that the scores of this group of participants with WS generally fell 'in the normal range', according to a standardised test of face recognition. Figure 1 depicts the Benton scores for the WS group plotted against individuals' chronological age, on the one hand, and against their Pattern Construction age equivalent, on the other. This figure includes the 15 of the 17 participants for whom Benton scores were

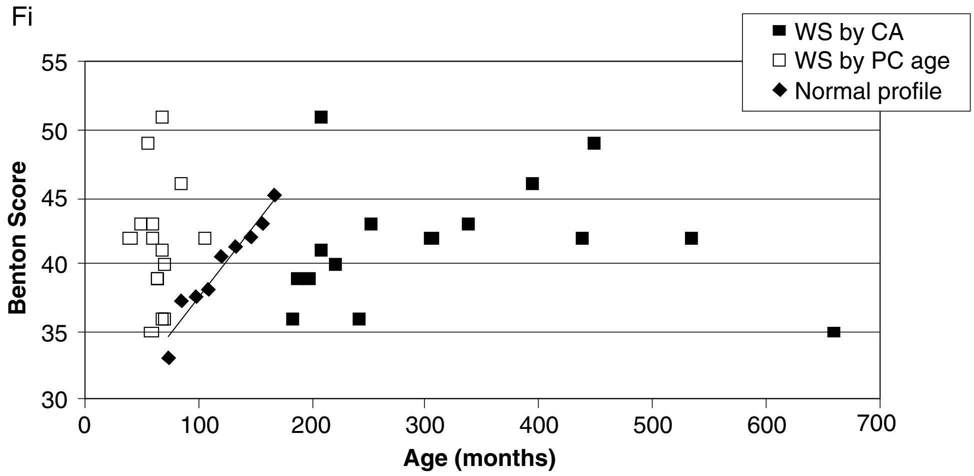

Figure 1 Benton face recognition scores for participants with WS ($N = 15$), plotted either against Chronological Age (CA) or against Pattern Construction (PC) age equivalent score. Also shown are normative scores from Benton age equivalent (Benton et al., 1983)

available. Scores on the Benton were also converted into a Benton age equivalent (Benton et al., 1983) for later use. Several points are of note. First, as in previous studies, many of the WS group fell within the normal range on the Benton task and, indeed, the overall WS group mean was in the normal range. Second, WS performance was nevertheless *delayed* compared to the typical developmental profile, so that scores-in-the-normal-range do not imply here entirely normal development. Third, while the WS scores broadly increased with age, this relationship was not significant ($r^2 = .000$, $p = .954$, or with 2 outliers removed, $r^2 = .153$, $p = .167$). This was because of the cross-sectional nature of the sample: severity of expression of the disorder varies independently of age. Fourth, the Benton scores were in excess of the level that would be predicted by the Pattern Construction ability of the WS group. Lastly, there was no correlation between performance on Benton and test age on the Pattern Construction subtest in the WS group ($r^2 = .000$, $p = .967$), suggesting that these two tasks tap (at least in part) different processes. While Brace et al. (2001) had no Benton and Pattern Construction data for their children, we collected indicative data in our laboratory for a sample of 21 healthy children between 3;6 and 11;2 (mean 6;9, SD 2;7) who performed both these tests. For these children, Pattern Construction test age strongly predicted Benton scores, with a correlation of $r = .873$ ($r^2 = .763$, $p < .001$).

The normal developmental trajectory. Trajectories were analysed using an analysis of co-variance (ANCOVA). This test requires the relationship between performance and age to be roughly linear. To linearise the relationships, accuracy was plotted against one-over-age squared (in months) (see Thomas et al., 2001, for details). The log of reaction time was plotted against the log of age (in months).

For the control sample, accuracy rates on upright and inverted faces (each out of 3) were compared to chronological age. Controls exhibited a significant improvement in accuracy with age ($F(1,109) = 27.01$, $p < .0001$) and a significant accuracy cost of recognising inverted faces ($F(1, 109) = 4.56$, $p = .035$). The size of the inversion effect did not alter significantly across the developmental profile ($F(1,109) = .153$, $p = .697$). Figure 2(a) depicts this relationship (note that, for clarity, accuracy is plotted against chronological age in this figure, rather than the transformed age variable used in the analysis). Chance performance in this face identification task was 11%.

Mean reaction times for upright and inverted face recognition on correct trials were compared to chronological age. Controls exhibited a significant reduction in reaction time with age ($F(1,109) = 38.50$, $p < .001$) and a significant time cost of recognising inverted faces ($F(1,109) = 4.50$, $p = .036$). Moreover, the cost of recognising inverted faces significantly increased with age, consistent with the emergence of configural face-processing expertise ($F(1,109) = 5.93$, $p = .016$). This trajectory is depicted in Figure 3(a).

WS developmental trajectory. For the relationship between performance and chronological age, direct comparisons between the WS and control groups must be interpreted with caution, because the samples have differing variability and are non-overlapping in terms of CA. The direct comparison revealed no overall significant group difference in either accuracy or reaction time when the performance of the WS group was compared to this much younger typically developing control group.

When considered on its own, the WS group revealed no significant relationship between accuracy or reaction time with increasing chronological age (accuracy: $F(1,12) = 1.13$, $p = .309$; RT: $F(1,12) = 2.33$, $p = .153$). Unlike the significance levels in the controls, there was only a trend in the direction of an inversion cost in the accuracy data ($F(1,12) = 4.34$, $p = .059$), and no effect in the RT data ($F(1,12) = .00$, $p = .999$). There was no indication that the inversion effect altered across chronological age (interaction of orientation and age: accuracy, $F(1,12) = 1.33$, $p = .271$; RT: $F(1,12) = .01$, $p = .908$). These trajectories are shown in Figures 2(b) and 3(b).

The next step was to explore whether any of the standardised measures – Benton, Pattern Construction,

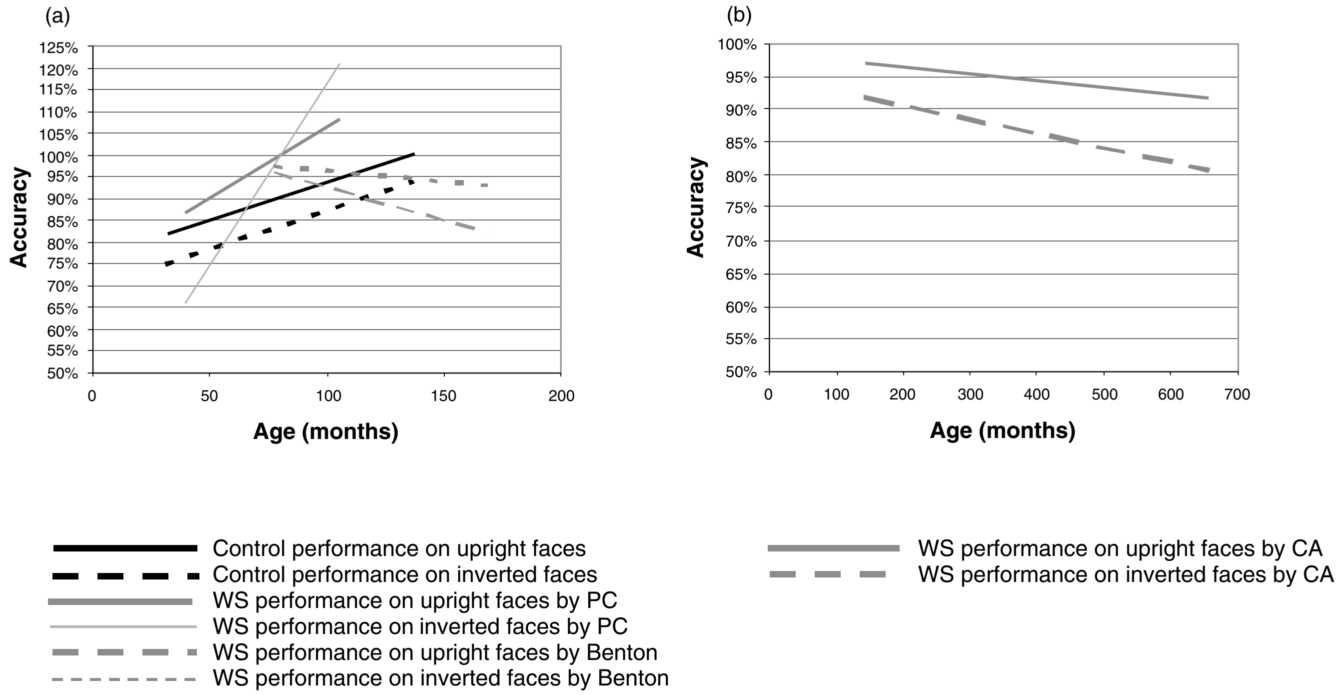

Figure 2 Developmental trajectories for accuracy on the Picture-book face recognition task: (a) trajectories for accuracy against age for controls and for the WS group plotted according to their Pattern Construction age equivalent and separately according to their Benton age equivalent; (b) accuracy against chronological age for the WS group

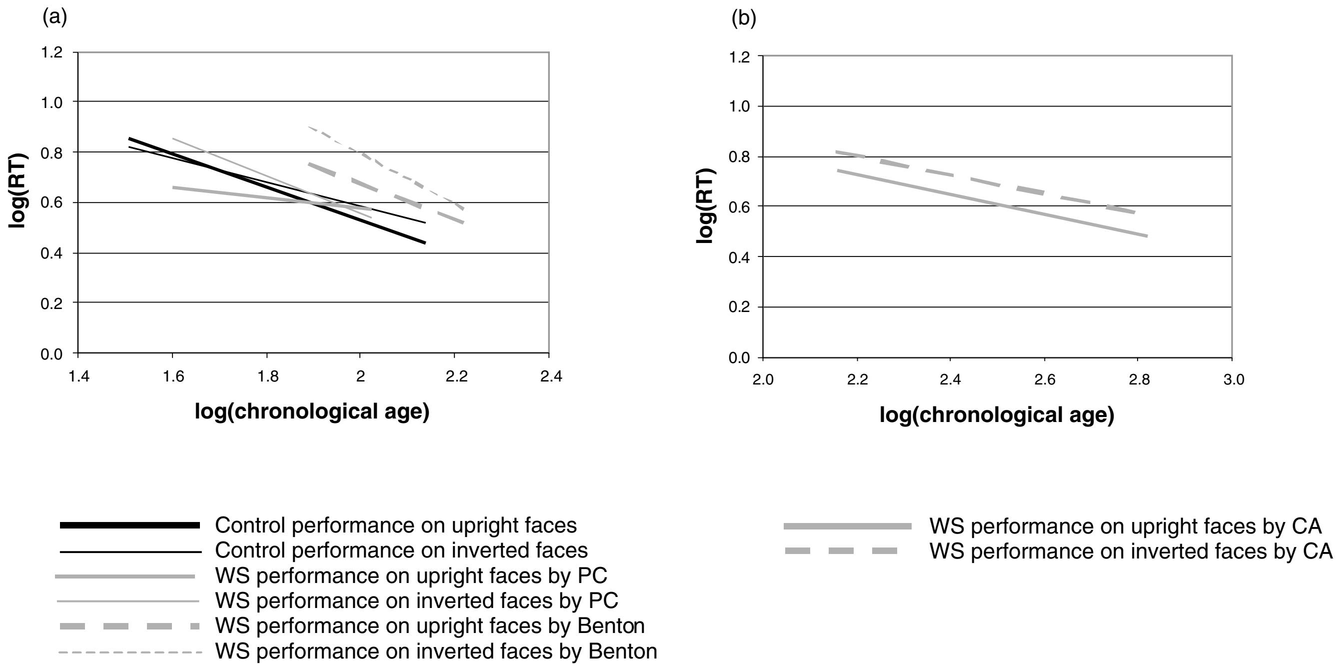

Figure 3 Developmental trajectories for response times on the Picture-book face recognition task: (a) response time against age for controls and for the WS group plotted according to their Pattern Construction age equivalent and separately according to their Benton age equivalent (log-log plot); (b) response time against chronological age for the WS group (log-log plot)

or BPVS – was a good predictor of performance on the Storybook task. Performance was compared against WS test age equivalents for each of these measures.

Performance on the Benton did not predict accuracy ($F(1,11) = .62$, $p = .447$) but did successfully predict reaction time on the Storybook task ($F(1,11) = 5.26$, $p = .043$; inversion effect non-significant: $F(1,12) = 3.23$, $p = .100$). By contrast, performance on the BPVS predicted neither accuracy nor reaction time. Performance on the Pattern Construction sub-test of the BAS-II did not predict reaction times but did successfully predict accuracy levels in the WS group (accuracy:

$F(1,12) = 10.33$, $p = .007$; RT: $F(1,12) = .90$, $p = .362$). The trajectories are shown in Figures 2(a) and 3(a). The accuracy trajectory revealed an unusual pattern in the WS group. There was a marginally non-significant inversion effect in the WS group ($F(1,12) = 4.54$, $p = .055$) but a significant interaction whereby the inversion effect became *smaller* as Pattern Construction ability increased ($F(1,12) = 12.91$, $p = .004$). The reaction time data replicated this pattern but differences were not reliable (inversion effect: $F(1,12) = 1.32$, $p = .273$; interaction: $F(1,12) = 1.13$, $p = .308$).

Comparison to the control group revealed that WS accuracy on the Storybook task was in excess of what would be expected for Pattern Construction ability (main effect of group: $F(1,121) = 4.21$, $p = .042$). The differential modification of the inversion effect with increasing age was not significant in the combined analysis (3-way interaction of group by age by orientation: $F(1,121) = 2.78$, $p = .098$).

Discussion

As in previous studies, we found that WS Benton scores fell within normal range and better than would be expected by their performance on a standardised visuospatial constructive task. Experiment 2 also yielded novel findings, because we built a full trajectory of typical developmental changes over time for this particular task. Indeed, between the ages of 3 and 12, the normal developmental trajectory reveals increasing accuracy and faster response times for upright faces, as well as the progressive emergence of an inversion effect in terms of reaction times. Most theorists concur that this is the signature of an emerging specialisation for configural processing of upright faces. While the WS group did not differ from the control trajectory in terms of accuracy or speed of response, their developmental trajectory failed to show the progressive emergence of a face inversion effect. In other words, as was suggestive of the findings from Experiment 1, the WS *behavioural* proficiency on some face-processing tasks, e.g., the Benton, seems to stem from an atypical developmental trajectory. In terms of relating WS performance to their other skills, increasing ability on the Benton task predicted an increase in reaction time in the WS group, but again, no emergence of an inversion effect. Benton performance did not significantly predict accuracy on the Storybook task. Ability on BPVS bore no relation to increasing performance on our task. Increasing ability on Pattern Construction predicted an increase in accuracy in the WS group and a modification of the inversion effect, but the modification was unusually in the reverse direction: better Pattern Construction predicted a smaller inversion effect.

Experiment 3: Configural and featural processing of schematic stimuli

The aim of this experiment was to assess whether adults with WS show a configural or featural bias in their perceptual processing of schematic faces and geometric patterns. The use of schematic faces allows tighter control of stimulus attributes and the closer matching of face and non-face stimuli. Participants once again were required to make similarity judgements about transformed stimuli. In one condition, they judged which of two test patterns differed most from a previously presented target pattern, where one test pattern was identical to the target and one was transformed. This task is similar to the *Difference Detection* condition in Experiment 1. In a second condition, both test stimuli were transformed, one configurally, one featurally. This *Judgement Preference* task allowed us to assess the relative salience of the two types of transformation for the participants. In addition, the coexistence of normally developing performance on Difference Detection but atypical performance on Judgement Preference would point to a preferred processing strategy in WS rather than an underlying deficit.

Method

Participants. Twelve adolescents and adults with WS were tested on this task. Their mean CA was 27;1 (SD: 11;11, range: 15;1–52;3). Language ability was assessed using the British Picture Vocabulary Scale (Dunn et al., 1997), for which the mean test age was 12;2 (SD: 3;10, range: 7;0–17;6). Visuospatial ability was assessed using the Pattern Construction subtest of the British Abilities Scale. The mean spatial test age of the WS group was 5;8 (SD: 1;3, range 4;1–8;9). Thus as in the previous experiments, the new WS group exhibited the characteristic disparity between these two abilities (language greater than visuospatial skill: paired t-test, $t = 6.88$, df $= 11$, $p < .001$), along with overall delay. Performance on the Benton for this group yielded a mean of 42.0 (SD: 3.4, range: 35–47) corresponding to performance within the normal adult range (41–54; Benton, Sivan, Hamsher KdeS., Varney, & Spreen, 1994). Once more, the age norms from the Benton were used to convert these scores into age equivalents, taking a ceiling score to be reached at 14 years. Employing this method, the WS group had a mean Benton age equivalent of 11;3 (SD: 2;5, range: 6;6–14;0). For comparison with Experiment 1, once more Pattern Construction test age failed to reliably predict Benton score in this WS sample, Pearson correlation = .264 ($r^2 = .070$, $p = .408$).

Sixty-one control children, adolescents and adults, covering the mental and chronological age spans of the clinical group, were also tested in order to build a developmental trajectory specifically for this task. Their overall mean CA was 12;5 (SD: 9;3, range: 5;5–53;1). There were 12 children between 5;4 and 6;5, 12 children between 7;0 and 7;5, 12 children between 8;10 and 9;9, 13 children between 12;2 and 12;10, and 12 adolescent and adults between 14;11 and 53;1. For the children, BPVS data were also collected to ensure that they fell within the normal range.

Stimuli. Three sets of stimuli were displayed on a computer screen: schematic faces, scrambled faces and geometric shapes. The stimuli all comprised four black elements within a yellow circle of 7 cm. Two featurally modified and two configurally modified versions of each stimulus were created. For the schematic and scrambled faces, featural changes were made by replacing the eyes, since this is the most salient feature of a face: round eyes were replaced with squares or diamonds of a similar size as the original feature. By contrast, since all features have equivalent salience for geometric patterns, all four features were replaced to avoid disrupting symmetry. Configural changes were made by stretching or squashing the set of features towards or away from the centre by 20 pixels, thereby creating new, second-order configural relations between the features. Configural changes to the scrambled face could not be made by displacing the elements vertically without breaking the symmetry of the arrangement, so they were made by displacing all the elements horizontally towards or away from the centre of the pattern by 20 pixels.

Procedure. In each trial, a target pattern was presented followed by, or simultaneously with, two test patterns. In the *Difference Detection* task, one of the test patterns was identical to the target pattern while the other was a version that had undergone either a featural or a configural transformation. In the *Judgement Preference* task, both of the test patterns were transformed versions of the target pattern: one featurally transformed, the other configurally transformed. For both tasks, the participant had to decide as quickly as possible which test pattern (left or right) differed the most from the target pattern.

The stimuli consisted of two blocks of 48 trials. One block had sequential presentation, with the target presented in isolation, to be replaced by the two test patterns. The second block had simultaneous presentation, with target and two test patterns appearing on the screen at the same time. There were 16 schematic faces, 16 scrambled and 16 geometric pattern trials. The order of these blocks was the same for all participants. Within each of these blocks there were 8 difference detection and 8 judgement preference (4 configural and 4 featural) trials, which appeared in a randomised order. Participants indicated which image they thought was most different by pressing the response key on the same side as the picture. Velcro pads were attached to the two relevant keys on the computer keyboard to assist participants with remembering which keys should be pressed. Each participant was tested individually and given 8 practice trials. For the Difference Detection trials the dependent variables were accuracy of response and response time, whereas for Judgement Preference trials it was the number of featural responses.

Results

As in the previous experiment, the results were analysed by building and comparing task-specific cross-sectional developmental trajectories for the control group and the WS group. For the WS group, the trajectory linking age and performance employed a variety of age measures: either their CA, their Benton age equivalent, their Pattern Construction age equivalent, or their BPVS age equivalent. As before, analysis of covariance was used to compare developmental trajectories. Unless otherwise noted, these involved the same data transformations to improve linearity discussed for Experiment 2.

The typical developmental trajectory. A comparison of control accuracy levels for all stimuli (face-like, scrambled, geometric), transformation types (featural vs. configural), and presentation conditions (simultaneous vs. sequential) against age was carried out. Presentation condition had no main effect ($F(1,58) = .21$, $p = .648$) and the same held for an overall analysis of response times ($F(1,58) = 1.07$, $p = .305$). There did appear to be a face-specific effect of simultaneous presentation, a point to which we will return shortly. For the remainder of the analyses, performance was averaged over simultaneous and sequential presentation conditions. A comparison of the two non-face-like patterns revealed that scrambled face features were harder than geometric patterns (accuracy: $F(1,58) = 3.96$, $p = .051$; RT: $F(1,58) = 1.92$, $p = .171$) but there were no interactions with other variables. For the remainder of the analyses, we will focus on comparisons between face-like stimuli and geometric patterns.

Difference Detection. The relationship between accuracy and age for the typically developing controls on face-like stimuli is depicted in Figure 4(a). An ANCOVA revealed significant improvement in accuracy with age ($F(1,59) = 4.88$, $p = .031$). There was no differential response to face-like stimuli and geometric patterns ($F(1,59) = .74$, $p = .392$), although there was an indication that performance on geometric patterns increased more rapidly with age (interaction of stimulus type and age: $F(1,59) = 3.95$, $p = .051$). Transformation type had no overall effect, but this turned out to be a consequence of averaging across presentation conditions.

A comparison of the accuracy data in the simultaneous vs. sequentially presented conditions revealed one unexpected effect. In the simultaneous condition, the controls responded more accurately to configurally transformed stimuli than featurally transformed stimuli. This held only for face-like stimuli and not for geometric patterns ($F(1,58) = 4.21$, $p = .045$) and was not present in the sequential conditions (interaction with condition: $F(1,58) = 3.96$, $p = .050$). An analysis of the sequential condition on its own revealed an interaction of transformation type with age, such that detection of configural transformations was less accurate than featural transformations, but the disparity decreased with age (interaction of transformation and age: $F(1,59) = 6.37$, $p = .014$). This effect was the same for both faces and geometric patterns. It is therefore possible that under the low-memory load conditions of simultaneous presentation, configural expertise for face processing emerges more easily in typically developing children than for sequential processing.

An analysis of reaction times revealed that responses became faster with age ($F(1,59) = 47.36$, $p < .001$), shown in Figure 4(b). Configural transformations were detected significantly more slowly than featural transformations but this disparity disappeared with age (main effect of transformation: $F(1,59) = 9.96$, $p = .003$; interaction with age: $F(1,59) = 6.90$, $p = .011$).

Judgement Preference. Which transformation type was more salient to typically developing children? Comparison of preference data was restricted to accuracy, because when participants' preferences were exclusively for one transformation type, no response

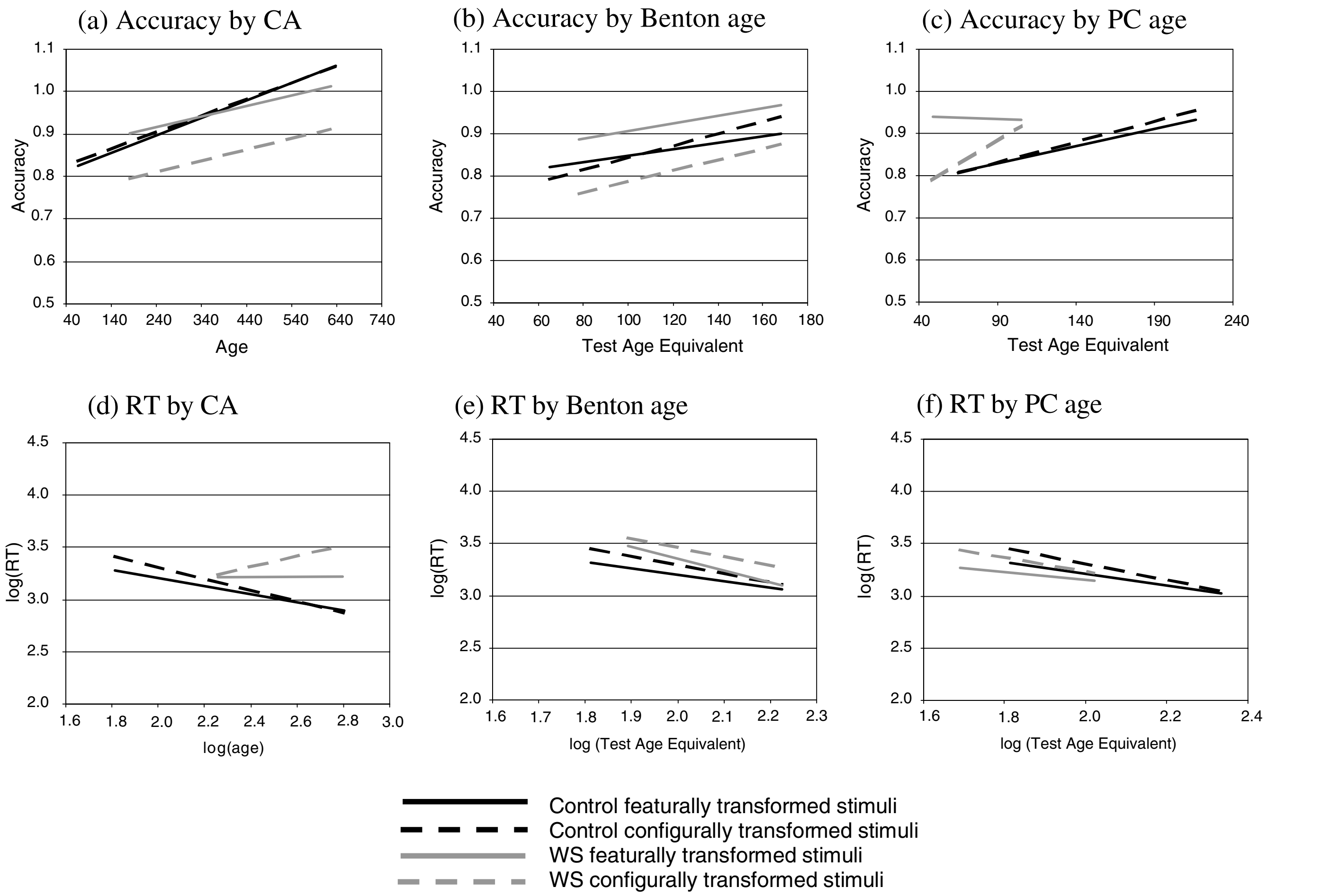

Figure 4 Cross-sectional developmental trajectories for the WS and control groups on the schematic faces task, for detecting featural vs. configural transformations. Trajectories for accuracy and reaction time across age are plotted according to the chronological age of the participants (CA), Benton face recognition test age equivalent (ceiling age 14;0), or Pattern Construction (PC) test age equivalent (ceiling age 18;0). See text for details

times were available for the non-preferred transformations. Figure 5(a) demonstrates how preferences change with age. The relationship between choice type and age was best fitted by a linear relationship, and so choice was compared against untransformed chronological age. The analysis was carried out for a single choice type, since configural vs. featural choices co-specify each other in the forced choice paradigm. An ANCOVA revealed a significant effect of age on choice type ($F(1,59) = 11.35$, $p = .001$) whereby in typical development, individuals increasingly see configurally transformed stimuli as more different from the target and decreasingly choose featurally transformed stimuli. Let us now turn to the trajectory of the clinical group.

The WS developmental trajectory. Difference Detection. Did the WS group demonstrate a normal facility for recognising each transformation type? Dealing with accuracy first, a between-group comparison of WS developmental trajectory with the typical trajectory revealed no overall effect of group ($F(1,69) = .216$, $p = .643$). The trajectories are shown in Figure 4(a). Importantly, however, the WS group exhibited worse performance on configurally transformed stimuli while the control group did not (interaction of group and transformation type: $F(1,69) = 8.27$, $p = .005$). Moreover, when the trajectory of the WS group was constructed according to their Benton test age equivalent, a comparison with controls revealed that the disparity in configural processing in WS persisted ($F(1,69 = 8.07$, $p = .006$), depicted in Fig. 4(b). In neither case is this pattern modified by presentation condition. Importantly, then, for their level of performance on a standardised face recognition task (which generally fell in the normal adult range), the WS group exhibited a differential pattern of response to featural vs. configural transformations.

A further comparison was carried out constructing the WS trajectory according to their test age on the Pattern Construction sub-test of the BAS (Elliot et al., 1996), depicted in Figure 4(c). In this case, the configural deficit now disappears ($F(1,69) = .21$, $p = .652$). The only significant difference between the trajectories is that the WS group shows a larger initial deficit in and subsequent steeper improvement for the geometric patterns than for faces, causing an interaction of group, stimulus type, and age ($F(1,69) = 5.56$, $p = .021$).

In the control group, we saw high accuracy on configurally transformed face-like stimuli in the simultaneous presentation condition alone. Examination of the WS performance in this condition yielded no such face-specific effect: configural transformation detection

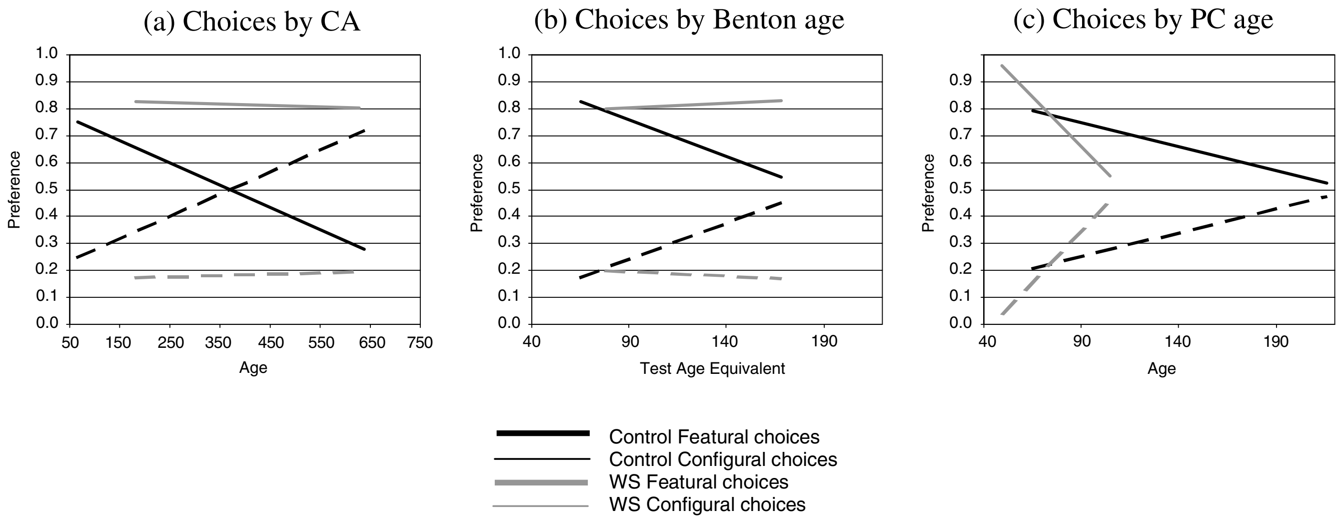

Figure 5 Cross-sectional developmental trajectories for the WS and control groups on preferences for featurally vs. configurally transformed face schematics, in a forced choice task to determine which stimulus is most different from the target. Trajectories are plotted according to the (a) chronological age of the participants (CA), (b) Benton face recognition test age equivalent (ceiling age 14;0), or (c) Pattern Construction (PC) test age equivalent (ceiling age 18;0). See text for details

lagged behind featural, as it did with geometrical patterns (effect of transformation: $F(1,10) = 5.87$, $p = .036$; interaction with stimulus type: $F(1,10) = .25$, $p = .632$). Moreover, the effect also held whether accuracy was plotted against CA or Benton test age equivalent.

An analysis of response times against age in the WS group generated the same configural deficit found in the accuracy data (interaction of transformation type and group: $F(1,69) = 5.79$, $p = .019$), in Figure 4(d). WS responses were slower overall, and did not show the decrease in RT with age of the control trajectory (main effect of group: $F(1,69) = 8.06$, $p = .006$; interaction with age: $F(1,69) = 10.09$, $p = .002$). However, these disparities became non-significant when RTs were plotted according to Benton test age, depicted in Figure 4(e). Plotting the WS trajectory according to Pattern Construction ability also caused the WS group to be indistinguishable from the normal trajectory, in Figure 4(f). However, somewhat unexpectedly, plotting performance against BPVS age equivalent produced a strong interaction with transformation type, such that increasing vocabulary age in the WS group (but not the control group) was associated with a divergence in RTs for recognising featurally and configurally transformed stimuli, with configural recognition slowing ($F(1,69) = 10.85$, $p = .002$). While this pattern appeared in the response times for both face-like stimuli and geometrical patterns, it did not appear in the accuracy data.

Judgement Preference. Did the WS group reveal the same relative salience of transformation type as the typically developing controls? Figures 5(a) to (c) summarise the responses of the WS group in deciding whether a configurally or featurally transformed stimulus was more different than the target. These are plotted against CA, Benton test age, and Pattern Construction test age respectively. The WS group demonstrated no alteration in choice with CA ($F(1,10) = .02$, $p = .880$) or Benton test age ($F(1,10) = .42$, $p = .534$; controls: $F(1,59) = 8.31$, $p = .005$). Plotting the WS performance according level of Pattern Construction ability produces a more typical-looking pattern, although the effect of age on choice does not yet approach significance ($F(1,10) = 1.91$, $p = .197$). The implication here is that sensitivity to the configural–featural dimension in WS is in line with Pattern Construction ability; yet this dimension is not the one exploited when participants with WS achieve their high face recognition scores in standardised tasks like the Benton.

Discussion

Although in Experiment 1 scores were poorer in the configural than in the featural trials in the upright condition, in this third experiment configural processing turned out to be easier than featural processing in upright condition. Why is this? A possible explanation is that the transformation types had different relative salience in the two tasks, i.e., configural transformations were more obvious with the schematic stimuli (stretched or compressed faces) than featural transformations (diamonds instead of squares for eyes), whereas in the Jane real faces, the configural changes involve quite subtle rearrangements compared to the more obvious change of different eyes in a face. Hence the importance of avoiding premature conclusions from a single study. Our focus, however, is on change over time in the sensitivity to the two transformation types, and particularly in the cross-group comparison, leading to several important differences between the WS group and the controls. First, the clinical group display a developmental delay in configural processing for their CA, but importantly, this also holds with respect to their level of face recognition performance on the Benton, which suggests atypical underlying processes. Second, when memory load is completely reduced due to simultaneous presentation of the target and test stimuli to be compared, then the controls, but *not* the participants with WS, display a face-specific sensitivity to configural transformations. Third, the clinical population is significantly slower than the controls, with a particular deficit in RTs for configurally transformed stimuli. While this holds for chronological age, the trajectory of RTs is not distinguishable from the normal trajectory when plotted according to their Benton or Pattern Construction age equivalents. Plotting the WS data according to their BPVS language test age does point to some indication of verbally mediated strategies that have been identified in other areas like number (Ansari et al., 2003), suggesting that the WS individuals may use their language proficiency to bootstrap other domains.

General discussion

What have we learnt from these three experiments? In our first experiment with real faces, we hypothesised that if face processing in WS had developed normally, then individuals with the syndrome should show no difference in accuracy or response time in discriminating a target face from a featurally or configurally transformed version of the target. At first blush, our overall results suggest this to be the case: the clinical group was as accurate as the controls on both identity recognition and difference detection, albeit somewhat slower. However, a key condition separated the groups' performance; for upright configural faces, the controls were significantly more accurate than the WS group. Yet, sensitivity to second-order configural differences in upright faces is the hallmark of face-processing expertise (Yin, 1969), lacking in the WS group. In our second experiment with real faces embedded in a story, we hypothesised that if face processing in WS had developed normally, then like the controls they would become increasingly accurate and rapid with age with respect to upright faces, but also show increased sensitivity to inversion, which disrupts configural processing. Our results showed that although again the WS group did not differ from the controls in terms of accuracy or speed of response, their developmental trajectory failed to show progressive emergence of the face inversion effect, again lacking the hallmark of face-processing expertise (Yin, 1969). Moreover, despite normal scores on the Benton, these latter did not predict accuracy on the storybook task. Finally, in our third experiment with schematic faces, we hypothesised that if face processing in WS had developed normally, then like controls they would demonstrate an emerging skill in detecting configural transformations in upright faces with increasing age. Our results showed not only a delay in configural processing with respect to CA, but importantly with respect to their level of face-processing performance on the Benton, suggesting not only delay but atypicality. Moreover, when memory load was reduced in the simultaneous presentation condition, the controls, but *not* the participants with WS, showed an increase in their sensitivity to configural transformations in faces.

A number of our findings, particularly the lack of progressive emergence of the inversion effect, only became obvious from the comparison of the developmental trajectories. Plotting the WS data with respect to their levels over time on several other developmental criteria (standardised measures of face processing, pattern construction or language) brought forth differences than were not detectible in accuracy or speed alone. Comparisons of behavioural scores at specific ages (CA or MA) often suggested no difference between WS and controls. But these are static comparisons that do not elucidate the trajectory by which people reach their proficiency. So, in our view, there is a distinct advantage of building task-specific developmental trajectories. In general, many studies of developmental disorders fall within the theoretical framework of adult neuropsychology. In such an approach researchers ask whether an ability in an atypical group is 'intact' or 'impaired' and draw their inferences on the basis of whether the clinical group does or does not differ from normal controls matched, say, on mental age. Delay is often ignored or implicitly dismissed as irrelevant, with statements such as 'the clinical group did not differ significantly from the controls', omitting to recall in the discussion that the MA-controls were, say, 20 years younger than the atypical participants! Here and elsewhere we have argued that delay cannot be simply ignored because the dynamics of a developing system over time are vital influences on final outcome (Karmiloff-Smith, Scerif, & Ansari, 2003). But this doesn't merely hold for mental-age-matched controls. Even if a clinical population reaches the same scores as their chronological controls, it is still an open question as to whether they display those *behavioural* scores via the same *cognitive* processes as the controls and whether the trajectory by which they moved from childhood to adulthood followed the same *developmental trajectory* as controls (see discussion in Karmiloff-Smith, 1998). In other words, the vital issue for an in-depth comparison of two groups is whether the developmental trajectory shows the same pattern *over time*. Merely demonstrating the equivalence of two groups at Time Z tells us about Times X and Y and thus nothing of the route by which each group reached that level. Furthermore, even when Benton scores fell within the normal range, our trajectories showed that WS performance was nevertheless delayed compared to the typical trajectory, meaning that even 'scores-in-the-normal-range' does not imply entirely normal development. And in many cases, Benton scores for face processing in WS did not predict success levels on configural processing in our face-processing tasks.

We have argued in this paper that building full normal developmental trajectories of each specific task, and subsequently plotting the atypical trajectory on the trajectory, is a more informative way in which to address developmental questions. What we deem to be particularly interesting in our studies is that the broad pattern of scores in the end state was often not reliably distinguishable from chronological age, and yet the trajectory of development was atypical. In addition, by placing the clinical group at different points on the typically developmental trajectory of different standardised measures, we overcome the problem of having to recruit numerous different individually matched control groups for each and every standardised measure. In sum, our approach focuses on change over time rather than performance at a particular moment in development.

The important question for developmental disorders must always be, in our view, 'does it (face processing, language, number, and the like) *develop* normally or atypically'? We have tested this by asking: 'Is face recognition in WS in step with developmental markers established in the control sample?' While a truly longitudinal approach would obviously be ideal (but take some two decades!), our cross-sectional task-specific trajectories of a very large number of controls addresses this question more directly than groups matched statically on either mental or chronological age. For the practical purposes of the present set of experiments, and because of the claims in the literature that we were attempting to address, our youngest participants were 12 years of age. However, our lab is currently working on the featural/configural face-processing distinction in infants, toddlers and children, to complete the trajectories of both typically developing controls and those with developmental disorders. These are clearly very labour-intensive studies.

In conclusion, we agree that the extant literature has identified a proficiency with respect to holistic processing in WS (Deruelle et al., 2003; Tager-Flusberg et al., 2003). But, contrary to the claims of these authors, this work is not inconsistent with our claim that second-order configural processing is indeed impaired in WS in both face processing and general visuospatial processing. Our work measuring brain potentials (Grice et al., 2003) also points to particularly deficient integration of features into a configural whole in WS, suggesting that both face processing and visuospatial processing suffer from similar deficits. Yet, for two decades, the literature has claimed that face processing and visuospatial processing are independent of one another in WS, with the former 'spared' and the latter seriously impaired (Bellugi et al., 1988, 1994). Our findings suggest that this may not be the case. Although scores on the Benton were better than predicted by Pattern Construction scores, it was these latter that were more in line with our face-processing experimental tasks targeting directly the featural/configural distinction. And, whereas people with WS show improvements with age for the Benton, the same improvements are not apparent for tests where configural processing is essential. When configural processing is crucial, people with WS show a deficit (Deruelle et al., 1999), but when global/holistic processing is needed such as in the Navon task, then the WS resemble typically developing controls (Farran, Jarrold, & Gathercole, 2003). It is *second-order configural processing*, demonstrated by the lack of an emergent inversion effect, that yields an atypical developmental trajectory in this clinical population. In our view, spatial cognition is simply more vulnerable to second-order configural impairments in WS development, and many face-processing tasks can be solved by featural or holistic strategies, thereby camouflaging any configural deficit. In other words, while scores in one domain (face processing) may outstrip scores in the other domain (visuospatial cognition), both domains may be affected by similar deficient processes but one reveals this impairment more subtly than the other. Seeming dissociations in the outcome, then, do not necessarily entail dissociations all the way along the developmental pathway (Karmiloff-Smith, 1998).

In conclusion, we argue that the developmental approach taken here, highlighting the importance of building developmental trajectories, is an essential, additional methodology for uncovering the subtleties of the causes of developmental disorders in general, and of face processing in particular.

Acknowledgements

We should like to thank the Williams Syndrome Foundation, UK, for putting us in touch with families and particularly all the clinical and typically developing participants for their contribution. This research was supported by MRC Programme Grant No. G9715642 NIH Project Grant R21TW06761-01 to AK-S. We especially thank Daphne Maurer and Cathy Mondloch of MacMaster University, Canada, for making the Jane Faces Task available to us.

Correspondence to

A. Karmiloff-Smith, Neurocognitive Development Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK; Email: a.karmiloff-smith@ich.ucl.ac.uk

References

Ansari, D., Donlan, C., Thomas, M., Ewing, S., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology*, *85*, 50–62.

Bellugi, U.L., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). I. The neurocognitive profile of Williams Syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, *12*(Suppl 1), 7–29.

Bellugi, U., Lichtenberger, L., Mills, D., Galaburda, A., & Korenberg, J.R. (1999). Bridging cognition, brain and molecular genetics: Evidence from Williams syndrome. *Trends in Neurosciences*, *5*, 197–208.

Bellugi, U., Sabo, H., & Vaid, J. (1988). Spatial deficits in children with Williams Syndrome. In J. Stiles-Davis & M. Kritchevsky (Eds.), *Spatial cognition: Brain bases and development*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bellugi, U., Wang, P.P., & Jernigan, T.L. (1994). Williams syndrome: An unusual neuropsychological profile. In S.H. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 23–56). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Benton, A.L., Hamsher, K.S., Varney, N.R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: Tests*. New York: Oxford University Press.

Benton, A.L., Sivan, A.B., Hamsher K. deS., Varney, N., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Brace, N.A., Hole, G.J., Kemp, R., Pike, G., Duuren, M., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. *Perception*, *30*, 85–94.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Rectem, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familiar faces. *Brain and Cognition*, *2*, 257–284.

de Haan, M. (2001). The neuropsychology of face processing during infancy and childhood. In C.A. Nelson & M. Luciana (Eds.), *The handbook of developmental cognitive neuroscience* (pp. 381–398). Cambridge, MA: MIT Press.

de Renzi, E. (1986). Current issues in prosopagnosia. In H.D. Ellis, M.A. Jeeves, F. Newcombe, & A.W. Young (Eds.), *Aspects of face processing*. Dordrecht: Martinis Nijhoff.

Deruelle, C., Mancini, J., Livet, M., Cassé-Perrot, C., & de Schonen, S. (1999). Configural and local processing of faces in children with Williams syndrome. *Brain and Cognition*, *41*, 276–298.

Deruelle, C., Rondan, C., Mancini, J., & Livet, M. (2003). Exploring face processing in Williams syndrome. *Cognitie, Creier, Comportanent*, *7*, 157–171.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology General*, *115*, 107–117.

Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics: Seminars in Medical Genetics*, *97*, 164–171.

Dunn L.M., & Dunn L.M. (1981). *Peabody Picture Vocabulary Test–Revised*. Minnesota: American Guidance Service.

Dunn, L.M., Whetton, C., & Pintilie, D. (1997). *British Picture Vocabulary Scale*. Windsor, UK: NFER-Nelson.

Elliot, C.D., Smith, P., & McCulloch, K. (1996). *British Ability Scales*. NFER-Nelson, Windsor, UK.

Farah, M.J., Levinson, K.L., & Klein, K.L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia*, *33*, 661–674.

Farran, E.K., Jarrold, C., & Gathercole, S.E. (2003). Divided attention, selective attention and drawing: Processing preferences in Williams syndrome are dependent on the task administered. *Neuropsychologia*, *41*, 676–687.

Grice, S.J., de Haan, M., Halit, H., Johnson, M.H., Csibra, G., Grant, J., & Karmiloff-Smith, A. (2003). ERP abnormalities of visual perception in Williams syndrome. *NeuroReport*, *14*, 1773–1777.

Johnson, M.H., & de Haan, M. (2001). Developing cortical specialization for visual-cognitive function: The case of face recognition. In J.L. McClelland & R.S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives. Carnegie Mellon symposia on cognition* (pp. 253–270). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, M.H., & Morton, J. (1991). *Biology and cognitive development. The case of face recognition.* Oxford (UK) and Cambridge (USA): Blackwell.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge MA: MIT Press.

Karmiloff-Smith, A. (1997). Crucial differences between developmental cognitive neuroscience and adult neuropsychology. *Developmental Neuropsychology*, *13*, 513–524.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*, *2*, 389–398.

Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P., & Udwin, O. (1997). Language and Williams Syndrome: How intact is 'intact'? *Child Development*, *68*, 246–262.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double dissociations in developmental disorders? Theoretically misconceived, empirically dubious. *Cortex*, *39*, 161–163.

Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G., Paterson, S., & Karmiloff-Smith, A. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental Science*, *5*, 233–246.

Maurer, D., Le Grand, R., & Mondloch, C.J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*, 255–260.

Mervis, C.B., & Bertrand, J. (1997). Developmental relations between cognition and language: Evidence from Williams syndrome. In L.B. Adamson & M.A. Romski (Eds.), *Research on communication and language disorders: Contributions to theories of language development* (pp. 75–106). New York: Brookes.

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

Passarotti, A.M., Paul, B.M., Joseph, R., Bussiere, J.R., Buxton, R.B., Wong, E.C., & Stiles, J. (2003). The development of face and location processing: An fMRI study. *Developmental Science*, *6*, 100–117.

Rossen, M.L., Jones, W., Wang, P.P., & Klima, E.S. (1995). Face processing: Remarkable sparing in Williams syndrome. Special Issue, *Genetic Counseling*, *6*, 138–140.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development*, *10*, 59–65.

Tager-Flusberg, H., Plesa-Skwerer, D., Faja, S., & Joseph, R.M. (2003). People with Williams syndrome process faces holistically. *Cognition*, *89*, 11–24.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46*(2), 225–245.

Temple, C. (1997). *Developmental cognitive neuropsychology*. Hove: Psychology Press.

Thomas, M.S.C., Grant, J., Barham, Z., Gsödl, M., Laing, E., Lakusta, L., Tyler, L.K., Grice, S., Paterson, S., & Karmiloff-Smith, A. (2001). Past tense formation in Williams syndrome. *Language and Cognitive Processes*, *16*, 143–176.

Thomas, M.S.C., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioural and Brain Sciences*, *25*, 727–750.

Udwin, O., & Yule, W. (1991). A cognitive and behavioural phenotype in Williams syndrome. *Journal of Clinical and Experimental Neuropsychology*, *13*, 232–244.

Yin, R.K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Wang, P.P., Doherty, S., Rourke, S.B., & Bellugi, U. (1995). Unique profile of visuo-perceptual skills in a genetic syndrome. *Brain and Cognition*, *29*, 54–65.

Manuscript accepted 4 March 2004