COGNITIVE NEUROPSYCHOLOGY, 2010, 27 (8), 636–664

Psychology Press
Taylor & Francis Group

Face recognition impairments despite normal holistic processing and face space coding: Evidence from a case of developmental prosopagnosia

Tirta Susilo[1], Elinor McKone[1], Hugh Dennett[1], Hayley Darke[1], Romina Palermo[1], Ashleigh Hall[1], Madeleine Pidcock[1], Amy Dawel[1], Linda Jeffery[2], C. Ellie Wilson[3], and Gillian Rhodes[2]

[1]Department of Psychology, Australian National University, Canberra, ACT 0200, Australia
[2]School of Psychology, University of Western Australia, Perth, WA 6907, Australia
[3]Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, NSW 2109, Australia

Holistic processing and face space coding are widely considered primary perceptual mechanisms behind good face recognition. Here, however, we present the case of S.P., a developmental prosopagnosic who demonstrated severe impairments in face memory and face perception, yet showed normal holistic processing and face space coding. Across three composite experiments, S.P. showed normal-strength holistic processing for upright faces and no composite effect for inverted faces. Across five aftereffect experiments, S.P. showed normal-sized face aftereffects, which derived normally from *face* space rather than shape-generic mechanisms. The case of S.P. implies: (a) normal holistic processing and face space coding can be insufficient for good face recognition even when present in combination; and (b) the focus of recent literature on holistic processing and face space should be expanded to include other potential face processing mechanisms (e.g., part-based processing). Our article also highlights the importance of internal task reliability in drawing inferences from single-case studies.

Keywords: Face perception; Face recognition; Developmental prosopagnosia; Holistic processing; Face space; Composite effect; Face adaptation; Face aftereffect.

Correspondence should be addressed to Tirta Susilo, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA (E-mail: bagus.t.susilo@dartmouth.edu).

We are extremely grateful to S.P. for her generosity, time, and interest, and we thank Devin Bowles for referring her to us. We also thank: Bradley Duchaine and Raka Tavashmi for providing us with their Cambridge Car Memory Test (CCMT); Anne Aimola Davies for directing us to literature on 95% confidence intervals (CIs) in neuropsychology; Davide Rivolta and Mary Broughton for testing some controls on the Macquarie Centre for Cognitive Science Famous Face Test (MACCS FFT). Daphne Maurer and Mayu Nishimura codesigned (with L.J. and G.R.) the original method for the identity aftereffect task used in Experiment 4. This research was supported by Australian Research Council Grants DP0450636 and DP0984558 to E.M., DP0770923 to L.J., and DP0877379 to G.R. T.S. is grateful for scholarship support from Australian National University (ANU) Centre for Visual Sciences and overseas student fee waiver from ANU Department of Psychology. T.S. and E.M. designed the study, wrote the paper, and take responsibility for task selection, data analysis, results interpretation, and theoretical claims. T.S. tested S.P. R.P. contributed to choice of neuropsychological tests. All other contributions involved provision of previously unpublished tasks and/or control data: H. Dennett for task creation and control data for Experiment 6 (face expansion–contraction aftereffect) and Experiment 7 (horse aftereffect); H. Darke for task creation and control data for Experiment 2 (composite front-view adult faces); A.H. and M.P. for control data and its analysis for CCMT; A.D. for control data and its analysis for Experiment 1 (three-quarter-view composite); E.W. and R.P. for task creation and some control data for MACCS FFT 2008; L.J., G.R., R.P., and E.W. for task creation and control data for Experiment 4 (identity aftereffect).

http://www.psypress.com/cogneuropsychology http://dx.doi.org/10.1080/02643294.2011.613372

Developmental prosopagnosia (DP; also known as congenital prosopagnosia) is a lifelong deficit of face recognition that occurs despite normal intelligence, intact low-level vision, and absence of known acquired brain injury (e.g., Duchaine & Nakayama, 2006b). It is a heterogeneous disorder, in which the defining deficit of inability to remember facial identity is sometimes found in conjunction with deficits in other facial processing abilities (e.g., perception of various aspects of identity or expression) and sometimes not (e.g., Duchaine, 2011).

Here, we present a case of DP in which the affected individual shows a pattern of impaired and intact abilities that is of particular theoretical interest. We report "S.P.", a woman in whom we find that both holistic processing and face space coding are, as far as we can ascertain, completely normal. This of interest because holistic processing and face space coding have been the two perceptual mechanisms most commonly presumed in the literature to be associated with deficits in face recognition ability. S.P.'s results thus imply that, although these mechanisms might be *necessary* for successful face recognition, they can be *insufficient* to ensure good recognition, even when present in combination. A second contribution of our paper is methodological. We note that previous DP studies left out important control conditions, which we detail below. A final contribution of our paper is statistical. Previous DP studies that make inferences from single cases fail to take into account the internal reliability of their tasks (i.e., measurement error in the individual DP's task score). We show how this could render some previous single-case inferences unwarranted and demonstrate how future studies could take into account task reliability in their analysis.

Holistic processing in DP

Holistic processing is generally understood as an obligatory integration of visual information from across the entire region of upright faces (Maurer, Le Grand, & Mondloch, 2002; McKone & Yovel, 2009; Rossion, 2008; Tanaka & Farah, 1993). Its measures include the large reduction in face memory or discrimination for upside-down faces compared to upright faces (the "inversion effect"; Yin, 1969), the part–whole effect (which occurs for upright but not inverted faces; Tanaka & Farah, 1993), and the gaze-contingent window effect (again present upright but not inverted; Van Belle, de Graef, Verfaillie, Rossion, & Lefèvre, 2010). The most common measure, which we use here with S.P., is the *composite effect* (Young, Hellawell, & Hay, 1987): In this illusion, the appearance of the top half of one person's face alters when it is aligned with the bottom half of a different person (see Figure 1A), and the composite effect is measured as the disadvantage at identifying the top half in this aligned condition compared to a misaligned baseline.

The presumption that holistic processing is essential for face recognition ability (e.g., Maurer et al., 2002, but see Konar, Bennett, & Sekuler, 2010) predicts that individuals with prosopagnosia should always demonstrate impaired holistic processing. This hypothesis is supported in studies of acquired cases (patient G.G. in Busigny et al., 2010; patient P.S. in Ramon, Busigny, & Rossion, 2010). In DP studies, however, results are mixed. Some findings suggest impaired holistic processing: DP individuals have shown smaller inversion effects than controls (Behrmann, Avidan, Marotta, & Kimchi, 2005; Duchaine, Yovel, & Nakayama, 2007; Russell, Duchaine, & Nakayama, 2009), and two group studies have reported smaller mean composite effects than controls ($N = 14$, Avidan, Tanzer, & Behrmann, in press; $N = 12$, Palermo, Willis, et al., 2011). In contrast, three studies using individual-level analysis have reported normal composite effects in the majority of individuals (7 out of 8 cases, Le Grand et al., 2006; 4 out of 7, Schmalzl, Palermo, & Coltheart, 2008; $N = 1$, Williams, Berberovic, & Mattingley, 2007). These findings suggest heterogeneity within DP, with impaired holistic processing being a potential cause of the prosopagnosia in some cases, but not all.

Interpretations of the composite studies above suffer from a methodological limitation: none tested an inverted-face control. The inference that upright composite effects reflect holistic processing is valid only when participants show no or

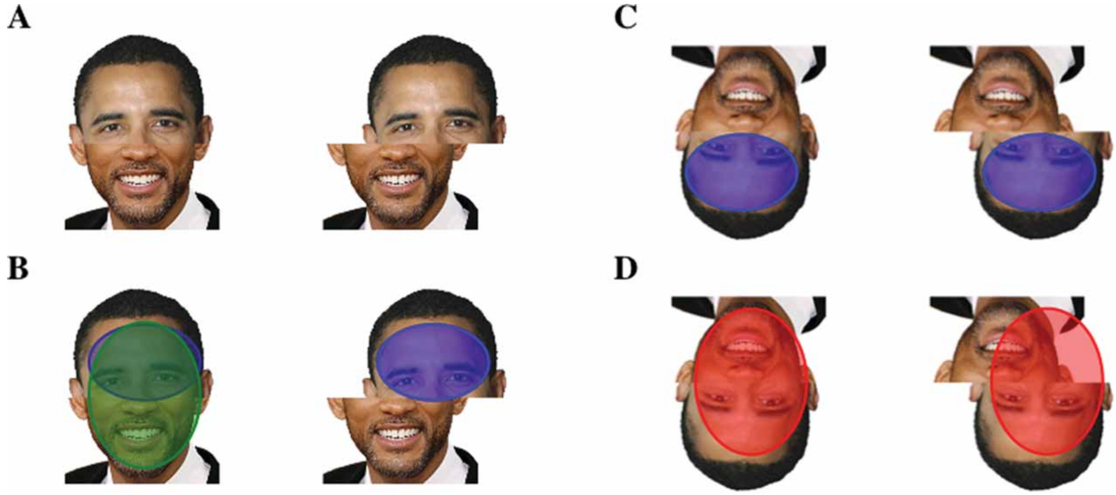

Figure 1. *Holistic processing as measured by the composite effect, and the importance of testing an inverted control. (A) In the upright orientation, two different identities (top-half Barack Obama, bottom-half Will Smith) integrate to generate an illusion of a third identity, when aligned but not when misaligned; this makes it slower and less accurate to identify the top half aligned than misaligned. The illusion disappears inverted (this can be seen by rotating the page). The composite effect is measured as the disadvantage aligned, as compared to misaligned. (B) Use of this composite measure implicitly assumes that spatial attention can always be narrowed solely to the target half (blue circle) except where holistic processing in the upright aligned condition forces mandatory processing of the entire face (green circle), and so (C) in the inverted orientation, spatial attention can be localized—producing no disadvantage for aligned halves—because holistic processing is absent. (D) If a prosopagnosic individual has spatial attention that is broader than typical (red circle), this will produce a false composite effect (i.e., present even in the inverted orientation) because more interfering information from the nontarget half falls in the spotlight of attention in the aligned condition than in the misaligned condition. Image in A taken from McKone, E., & Robbins, R. (2011). Are faces special? In A. J. Calder, G. Rhodes, M. H. Johnston, & J. V. Haxby (Eds.), Handbook of face perception. Oxford, UK: Oxford University Press (Fig. 9.2. p. 153). Copyright 2011 by permission of Oxford University Press (www.oup.com). To view a colour version of this figure, please see the online issue of the Journal.*

little composite effect for inverted faces. This is because the composite task rests on an implicit presumption that in the absence of holistic perception, participants are able to localize spatial attention to the target half of a composite face (see Figures 1B–1D for detailed explanation; McKone et al., 2010). In typical Caucasian participants, this presumption is supported by the standard findings of no or weak composite effects for inverted faces (Carey & Diamond, 1994; McKone, 2008; Robbins & McKone, 2003, 2007; Young et al., 1987). But this may not be true for other populations; for example, Asian participants seem to have a broader attention spotlight than Caucasians (e.g., a stronger global-over-local bias on the Navon task; McKone et al., 2010), and an unpublished study from our laboratory found a substantial composite effect for inverted faces in Asian participants that was not found in Caucasians (Wickramariyaratne, 2006). The atypical development of DP individuals raises the possibility that some may have problems in localizing spatial attention. Indeed, one Navon study reported a nonsignificant trend toward greater global bias in 14 DPs than in controls (Duchaine et al., 2007, although 5 DPs showed the opposite pattern in Berhmann et al., 2005). This argues for the importance of testing the inverted condition: Only where a DP shows no composite effect for inverted faces can we assert that their observed composite effect for upright reflects holistic processing.

Face space coding in DP

Face space coding refers to a multidimensional space in which individual faces are coded according to their values on a set of dimensions that best discriminate between those faces (Valentine, 1991).[1] Face space coding is commonly tapped via adaptation aftereffects (Webster & MacLin, 1999), in which prolonged viewing of a face temporarily shifts perception of the normal face away from the adaptor. These aftereffects occur in both "figural" versions (e.g., adapting to an eyes-up face makes an average eye height face appear eyes down), and an "identity" version (adapting to one person "Dan" makes the average morph face appear like his opposite on all facial attributes, i.e., "anti-Dan").

It has been suggested that face recognition deficits might be associated with abnormal face space coding: Reduced face aftereffects have been reported together with poor face recognition in autism spectrum disorder (Pellicano, Jeffery, Burr, & Rhodes, 2007) and in acquired prosopagnosia (patient A.P.; Nishimura, Doyle, Humphreys, & Behrmann, 2010). In contrast, DP individuals tested to date appear often to have a face space that is similar, at least in general structure, to that of normal controls. Le Grand et al. (2006) reported that some of their DPs had a normal pattern of rating an averaged face as more attractive than its component faces (i.e., the average-is-attractive effect). Nishimura et al. (2010) found that 6 DPs, on average, showed normal caricature effects and reasonably normal face dimensions determined from multidimensional scaling of face similarity ratings; their face aftereffects were not significantly smaller than those of controls, although note that these experiments did not change stimulus size between adaptor and test items and so did not control for potential low-level vision contributions to the aftereffects (Mayu Nishimura, personal communication, June 3, 2010).

We used face aftereffects to investigate face space coding in S.P. Face aftereffects allow us to go beyond testing whether a DP individual has a face space, and they can also be used to assess the coding sensitivity along the manipulated facial dimension/s—namely, the maximum accuracy with which the space can discriminate two faces differing by a unit amount along that trajectory. This idea derives from the evidence that face space uses opponent rather than multichannel coding (e.g., Freiwald et al., 2009; Robbins, McKone, & Edwards, 2007). In opponent coding (Figure 2), a facial dimension such as eye height is coded by the relative output of two oppositely tuned neural populations, one showing maximum response to very high eyes and minimum response to very low eyes, and the other the reverse. Adaptation shifts the physical face perceived as average because an adaptor placed away from the preadaptation average reduces the response of one pool more than that of the other (adaptation is proportional to initial response strength; Maddess, McCourt, Blakeslee, & Cunningham, 1988). An adaptor of fixed distance from the average (e.g., one with its eyes shifted up by "50 pixels") will produce an aftereffect that varies in magnitude depending on the steepness of an individual subject's underlying response functions (Figures 2A–2B). Specifically, the aftereffect size (i.e., the shift in the face perceived as most normal) will be larger for a subject with steeper tuning functions, which in turn correspond to greater sensitivity of face space's tuning along that dimension because a unit change in eye height becomes easier to detect (i.e., it gives a larger relative difference in the output of the two opponent pools).

Issues of task reliability in single-case inferences

Prosopagnosia researchers generally use tasks developed in the cognitive psychology tradition that are known to produce reliable results when scores are averaged across 15–25 participants. It is different, however, to claim that an *individual's* score is reliable: Unless a test has perfect internal reliability (i.e., split-half reliability or Cronbach's

[1] Note that little is known about the relationship between holistic processing and face space coding. It is not known whether one precedes the other in time course, or whether they occur in parallel, or even whether they might, in some currently not-understood manner, be two ways of tapping the same underlying mechanisms (for discussion, see Susilo, McKone, & Edwards, 2010a).

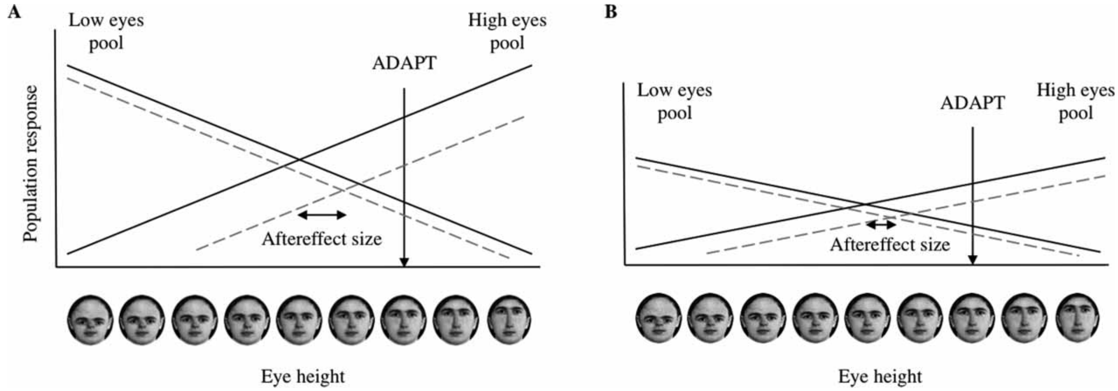

Figure 2. *The relationship between the size of face aftereffects and the steepness of neural tuning functions. Solid black lines indicate tuning responses prior to adaptation; dotted grey lines indicate tuning responses post adaptation. Adaptation to a face with high eyes will reduce activation of the high eyes pool more than activation of the low eyes pool, which results in the crossover point (i.e., the physical face stimulus perceived as having the most normal eye height) shifting to the right (i.e., the eye height aftereffect). Steeper tuning curves in (A) give better discrimination of a unit stimulus change and also produce greater shift (i.e., a larger face aftereffect) than shallower tuning curves in (B).*

alpha equal to 1.0) then there will always be measurement error in an individual case's score. (Note that this is a different, additional source of potential error than the effects of modest control sample size as discussed in Crawford & Howell, 1998.) Low internal reliability results in a wide confidence interval (CI) around the score, even assuming an infinitely large control sample size (Ley, 1972). This means that some of the single-case inferences made in previous studies may not be warranted because the tasks might not have good reliability (Behrmann et al., 2005; Le Grand et al., 2006; Nishimura et al., 2010; Schmalzl et al., 2008; Williams et al., 2007).

Table 1 illustrates the task reliability issue, using an example taken from Le Grand et al. (2006, see their Table 3). The authors inferred that 7 of their 8 DPs had normal holistic processing based on their individual composite face scores (only case E.N. reached their criterion of abnormality by falling in the poorest 5% of the population). This inference, however, assumes that the composite task is highly reliable—an assumption that, based on available data to date, seems rather unlikely. Zhu et al. (2010) reported a composite task reliability of .65; similarly, the three composite tasks used in the present study (see Section "Experiments 1–3: Holistic Processing") had reliability indices of .49, .65 and .75.[2] We therefore calculated the CI for Le Grand et al.'s DP's scores assuming task reliability of .65. Using the lower bound of the 95% CI, half of the cases (4 of the 8) would be categorized as having impaired holistic processing, but using the upper bound, none would be considered impaired.

THE PRESENT STUDY

As reviewed above, several studies of DP have examined either holistic processing (Avidan et al., 2011; Behrmann et al., 2005; Le Grand et al., 2006; Palermo, Willis, et al., 2011; Schmalzl et al., 2008; Williams et al., 2007) or face space coding (Nishimura et al., 2010). No studies to date have examined both types of processing in the same individual using appropriate single-case statistics.

Here we investigate both holistic processing and face space coding in prosopagnosic individual S.P., and we also address methodological limitations

[2]Herzmann, Danthir, Schacht, Sommer, and Wilhelm (2008) computed a reliability of .23 for their composite task, but we excluded this report because their task did not produce a composite effect.

Table 1. *Reevaluation of composite task results in Le Grand et al. (2006, Table 3)*

Case	*Score*	z	*Inference by* z	*95% CI of score*	*95% CI of* z	*Inference by lower* z	*Inference by upper* z
A.S.	29	0.1	Normal	[17.40, 40.60]	[−1.06, 1.26]	Normal	Normal
B.C.	15	−1.3	Normal	[3.40, 26.60]	[**−2.46**, −0.14]	Impaired	Normal
D.J.	12	−1.6	Normal	[0.40, 23.60]	[**−2.76**, −0.44]	Impaired	Normal
E.N.	4	**−2.4**	Impaired	[−7.60, 15.60]	[**−3.56**, **−**1.24]	Impaired	Normal
H.H.	38	1	Normal	[26.40, 49.60]	[−0.16, 2.16]	Normal	Normal
J.H.	19	−0.9	Normal	[7.40, 30.60]	[**−2.06**, 0.26]	Impaired	Normal
M.T.	27	−0.1	Normal	[15.40, 38.60]	[−1.26, 1.06]	Normal	Normal
N.M.	50	2.2	Normal	[38.40, 61.60]	[1.04, 3.36]	Normal	Normal

Note: CI = confidence interval. DP = developmental prosopagnosia. Only E.N. was originally categorized as having impaired holistic processing ("Inference by *z*" column). Taking into account task reliability alters interpretations. Assuming task reliability of .65, we calculated the 95% CIs around each case's composite score (Ley, 1972). Inferences based on lower bound *z* scores indicate that 4 DPs had impaired holistic processing (B.C., D.J., E.N., J.H.), while inferences based on the upper bound *z* scores indicate that all DPs had normal holistic processing. Bold numbers indicate *z* scores less than −2. Calculations use the Le Grand et al.'s (2006) control $M = 28$, $SD = 10$; as in Le Grand et al.'s original analysis, we have ignored the additional effects of modest control sample size (Crawford & Howell, 1998) in the calculations of 95% CIs.

present in previous studies. To assess holistic processing, we use the composite task. To rule out an atypical attentional account for the composite effect, we test both upright and inverted orientations. To assess face space coding, we use face aftereffects. To rule out potential contributions from low-level vision, we change the stimulus size between adapt and test conditions (e.g., Webster & MacLeod, 2011). We also, for the first time in DP, examine whether any observed face aftereffects are generated specifically by face-level representations as opposed to midlevel shape-generic mechanisms (e.g., coding of aspect ratio; Rhodes & Leopold, 2011, Susilo, McKone, & Edwards, 2010a). Finally, to address the issue of measurement error in S.P.'s scores, we report the reliability of all our composite and aftereffect tasks and take this information into account in our statistical analysis of whether S.P. is impaired or normal on a given task. We also use multiple experiments (three composite experiments, five aftereffect experiments) to allow for converging evidence and to compensate where reliability for a particular task is only modest.

Case description

S.P. (subject F21 in Bowles et al., 2009) is a female Caucasian Australian holding a Bachelor of Arts/Science degree in Psychology, currently working as a public servant. She was unaware of her face recognition deficits prior to participating in an experiment in our laboratory when she was 21. In a detailed follow-up self-report, she described classic signs of prosopagnosia. She reports trouble following movies, particularly ones with older male characters (who all have similar hairstyle). Hollywood actresses all look similar to her, except those with distinctive features (e.g., Julia Roberts's smiling lips). She reports recognizing people by nonfacial information such as hair, voice, height, and walking style, although she realizes that these strategies are not always effective. Her most useful cue is hair, and she finds it difficult to recognize women with headscarves. She also finds it difficult to recognize photographs of familiar people when they were younger.

S.P. believes her prosopagnosia has negatively affected her social interactions and made her socially awkward. She needs to make a conscious effort to remember people. She does not call people by name or greet them publicly unless she is certain of their identity. She reports having had a lot of trouble getting to know her friends at school, and remembering who said what to whom. She found many of her friends looked the same because her rural school had little ethnic diversity, and students wore school uniforms. As an adult, she prefers to mix with a small number of people she knows, rather than socialize in large group settings. S.P.'s difficulties in social settings, however, do not reflect an autism spectrum disorder: Her score on the Autism Spectrum Quotient (Baron-Cohen et al., 2001) was 15, which is in the 41st percentile (see Table 2) and well below the autism cut-off of 32. Regarding other skills, S.P. reports she is good at visualizing, remembering, and spelling words. She has good grades in statistics and formal logic. She reports problems with space perception: It took her a long time to learn how to drive, particularly to reverse park; and she finds spatial mathematics difficult (calculus and trigonometry).

Our classification of S.P.'s prosopagnosia as most likely developmental is due to (a) a lack of any known acquired brain injury, (b) an apparent lifelong deficit as revealed in her self-reports, (c) her lack of awareness of any problem prior to our formal testing (in contrast to acquired prosopagnosics, who are usually aware of a sudden change in their everyday face recognition ability), and (d) her reports that face recognition problems might run in her father's side of the family (e.g., her father finds movies hard to follow; both her father and grandfather preferred a small social circle). However, since we did not have the opportunity to scan S.P.'s brain, we cannot exclude the possibility that her prosopagnosia is acquired, for example due to an unsuspected stroke during development. S.P. participated in testing and interviews for approximately 17 hours over two years (November 2007 to December 2009).

Neuropsychological profile

Calculation of percentile rank and 95% confidence intervals

We use S.P.'s percentile score on each task, because the percentile score directly reflects how common (or uncommon) her raw score is in a normative population (Crawford, Garthwaite, & Slick, 2009). We assessed her as impaired when her percentile rank fell in the poorest 2% of the population, and as unambiguously normal where even the lower bound of her 95% confidence interval (CI) did not fall even in the poorest 5%.

We report two types of 95% CIs whenever possible. First, on all tasks, we present 95% CIs based on comparisons of S.P.'s raw scores to those of the population using the modified t test (Crawford & Howell, 1998). The percentile scores and CIs were calculated using SINGLIMS.EXE (Crawford & Garthwaite, 2002) except in one subtest (the learning phase of the Cambridge Face Memory Test) where control scores are highly skewed, and we used the relevant formula (Crawford et al., 2009). Second, we calculated 95% CIs around the z scores based on task reliability, irrespective of the control sample size (Ley, 1972). The reliability formula was

Table 2. *S.P.'s scores on neuropsychological tests and percentile rank in the population*

		Controls					
Test	*S.P.*	*M*	*SD*	*N*	*t*	*Population rank*	*95% CI of population rank*
Raven Advanced Matrices	9	9.40	1.78	506	−0.22	41.12	[37.72, 44.58]
Word memory	45	51.50	12.47	36	−0.51	30.52	[19.31, 43.27]
Birmingham Object Recognition Battery							
Length	26	26.90	1.60	39	−0.55	29.09	[18.47, 41.25]
Size	26	27.30	2.40	39	−0.53	29.79	[19.08, 41.99]
Orientation	24	24.80	2.60	39	−0.3	38.15	[26.53, 50.62]
Picture Naming	15	12.70	2.20	34	1.03	84.48	[73.26, 92.78]
Autism Spectrum Quotient	15	16.40	6.30	174	−0.22	41.24	[35.48, 47.15]

Note: "Population rank" refers to the percentage of the population who perform worse than S.P., here calculated based on the Crawford and Howell (1998) t score (i.e., taking into account the control sample size). "95% CI of population rank" refers to the 95% confidence interval around this rank.

Cronbach's alpha where there were no missing trials and where the scores of interest were from single conditions (e.g., accuracy in the Cambridge Face Memory Test). Otherwise, when the measure was reaction time (with missing trials due to errors and outlying reaction times, RTs), or when the score of interest was a difference score between two conditions (e.g., a difference between performance in the aligned and misaligned conditions in the composite task), we used the Spearman–Brown corrected split-half reliability based on an average of 50 different splits. We are not aware of any statistical methods that combine the effects of both task reliability and modest control group into a single 95% CI.

Intelligence, verbal memory, elementary vision, and basic-level object recognition

As shown in Table 2, S.P. performed in the normal range on neuropsychological screening tests. S.P.'s intelligence was at least average, based on (a) a normal score (41st percentile) on the Raven Advanced Matrices Test (Raven, Raven, & Court, 1991; control data from Bors & Stokes, 1998), and (b) her academic performance: Her entry to the Australian National University required school performance in the top 25% of those students completing a university-entrance accredited program of study, and she was further accepted into the psychology honours programme. S.P. demonstrated no general memory impairment, scoring in the 30th percentile on a test of word memory (Bowles et al., 2009). On the Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993), S.P. performed in the normal range on low-level vision tests (discrimination of length, size, and orientation of lines; 29th, 30th, and 38th percentiles, respectively), and she also demonstrated good recognition of objects at the basic level, scoring at the 84th percentile on picture naming.

Face memory (famous and novel stimuli)

S.P. was impaired on memory for both famous and novel faces. Memory for famous faces was assessed with the Macquarie Centre for Cognitive Science Famous Face Test 2008 (MACCS FFT 2008; Wilson, Palermo, Rivolta, Williams, & Schmalzl,

Figure 3. *An example stimulus in the Famous Face Task (Macquarie Centre for Cognitive Science Famous Face Test; MACCS FFT 2008), showing actress Cate Blanchett. To view a colour version of this figure, please see the online issue of the Journal.*

2008), comprising 20 front-view photographs of faces famous to Australians and 20 unfamiliar faces matched on sex and age, with neutral or smiling expressions and external face regions removed (Figure 3). Faces are presented one at a time until response. Participants first decide whether the face is familiar or unfamiliar: if familiar, they then identify the face by either naming the person or describing other detailed information. The person's name and associated semantic information is then presented, and participants indicate whether they know this person. The measure is the number of faces identified as a percentage of the number of people known. Control participants were 25 Caucasians (18–33 years, 17 females) from the Australian National University and Macquarie University communities. The reliability of this test could not be computed because there was significant variability in terms of the number of famous faces with which the participants were familiar.

As shown in Table 3, S.P. was impaired on the famous faces task: Her score was poorer than the 1st percentile. This was not due to lack of knowledge about the famous identities, because S.P. was familiar with 18 out of 20 of them.

Memory for novel faces was assessed with the Cambridge Face Memory Test (CFMT, Duchaine

Table 3. *S.P.'s scores on face and car tests and her percentile rank in the population*

Test	Reliability	S.P.	*Controls* M	SD	N	z	Population rank z	95% CI of population rank z	t	Population rank t	95% CI of population rank t
Faces											
CFMT upright	.88	36	54.95	8.71	118	−2.18	**1.46**	[**0.22**, 6.68]	−2.17	**1.62**	[**0.61**, *3.27*]
Stage 1: Learn	.48	14	17.69	0.71	118	−5.20	**0.00**	**[0.00, 0.00]**	−5.17	**0.00**	**[0.00, 0.00]**
Stage 2: Novel	.81	14	22.14	4.98	118	−1.63	5.16	[**0.64**, 21.77]	−1.63	5.31	[*2.81*, 8.74]
Stage 3: Noise	.69	8	15.11	4.07	118	−1.75	*4.01*	[**0.23**, 25.46]	−1.74	*4.23*	[*2.10*, 7.24]
Stages 2 + 3	.87	22	37.25	8.47	118	−1.80	*3.59*	[**0.60**, 13.79]	−1.79	*3.78*	[**1.82**, 6.60]
MACCS FFT	n/a	44.44	84.60	12.60	10	−3.19	**0.05**	n/a	−3.04	**0.70**	[**0.00**, 5.45]
CFPT upright	.73	94	31.44	9.10	29	−6.87	**0.00**	**[0.00, 0.00]**	−6.76	**0.00**	**[0.00, 0.00]**
CFPT inverted	.5	100	59.83	13.08	29	−3.07	**0.09**	[**0.00**, *4.55*]	−3.02	**0.27**	**[0.00, 1.44]**
Cars											
CCMT upright	.75	38	50.44	7.15	93	−1.74	*4.09*	[**0.33**, 22.36]	−1.73	*4.34*	[**1.97**, 7.86]
Stage 1: Learn	.53	11	15.73	1.91	93	−2.48	**0.66**	[**0.01**, 12.92]	−2.46	**0.78**	**[0.20, 1.95]**
Stage 2: Novel	.64	11	17.91	4.12	93	−1.68	*4.65*	[**0.22**, 30.85]	−1.67	*4.93*	[*2.32*, 8.70]
Stage 3: Noise	.55	16	16.80	3.19	93	−0.25	40.13	[5.82, 85.54]	−0.25	40.18	[32.40, 48.26]

Note: CFMT = Cambridge Face Memory Test. CFPT = Cambridge Face Perception Test. CCMT = Cambridge Car Memory Test. MACCS FFT = Macquarie Centre for Cognitive Science Famous Face Test. "Population rank *z*" and "population rank *t*" refer to the percentage of control population who perform worse than S.P. as calculated using *z* scores (i.e., ignoring control sample size) and *t* scores (i.e., taking into account control sample *N*). "95% CI on population rank *z*" uses internal task reliability to give the 95% confidence interval on S.P.'s percentile rank. "95% CI on population rank *t*" uses control sample *N* to give the 95% confidence interval on S.P.'s percentile rank. Bold numbers indicate where S.P. performed in the poorest 2% of the population; italics indicate where S.P. performed in the poorest 5% of the population. Note that all *z* scores, *t* scores, and ranks are calculated such that negative values and low ranks indicate performances poorer than the control mean (i.e., higher raw score for CFPT; lower for all other tasks).

& Nakayama, 2006a). The CFMT has been demonstrated to have stronger theoretical validity than older alternatives: It avoids hair and clothing cues, shows large inversion effects, is very weakly correlated with general intelligence and only moderately correlated with nonface visual memory, has reasonably good hit rates in diagnosing prosopagnosia, and has high task reliability (Bowles et al., 2009; Duchaine & Nakayama, 2006a; Wilmer et al., 2010). Participants learn six male target faces, each in three views. Each test trial is a forced choice of three, one of which was one of the target faces. In Stages 2 and 3 of the test, the target faces have different views, poses, and lighting to the images in the learning phase; noise is further added in Stage 3. Total number of trials is 72. The performance measure is number of correct trials (chance score is 24, higher score reflects better performance). S.P. was tested on the upright version of the CFMT. Control mean, standard deviation, and Cronbach's alpha (.87) were computed from 118 Australian and New Zealand Caucasians aged 18–30 years (Bowles et al., 2009).

As can be seen in Table 3, S.P. was severely impaired on the CFMT. She scored below the 2nd percentile.

Face perception

S.P. was also impaired in the Cambridge Face Perception Test (CFPT; Duchaine et al., 2007). This test uses simultaneous presentation of target and test faces, to reduce memory demands. Each trial presents a target face in three-quarter view above six test faces in front view. Each test face is a morph between the target face and one of six different other faces, and all six test faces contain different proportions of the target face (ranging from 88% to 28%). Participants are given one minute to sort the test faces based on their similarity to the target face. There are eight target faces, each presented upright and inverted, for a total of 16 trials. The measure is the deviation from the correct ordering (higher scores reflect worse performance, chance score is 93). Control participants were 29 young Australian and New Zealand Caucasians (18–30 years) who produced a Cronbach's alpha of .73 for upright, and .50 for inverted (Bowles et al., 2009).

As shown in Table 3, S.P. was extremely impaired for upright faces, scoring more than 6 standard deviations worse than the mean. S.P. showed little inversion effect on the CFPT, but this finding is hard to interpret because her upright performance was already at chance.

Within-class discrimination of objects

We examined S.P.'s ability to perform within-class discrimination of objects using the Cambridge Car Memory Test (CCMT, created by Bradley Duchaine & Raka Tavashmi (Dennett et al., 2011)). This test (Figure 4) has identical testing structure and procedure as does the CFMT. Participants learn six target cars, each seen in three views. Each test trial is a forced-choice of three, one of which was one of the target cars. In Stages 2 and 3, targets have different view and lighting from those of the learning phase. Total number of trials is 72 (chance score is 24, higher score reflects better performance). Control participants were 93 females (17–28 years) from the Australian National University community; only females were used as controls because the task produced a substantial sex difference on mean performance. Cronbach's alpha reliability for females was .75 (Dennett et al., 2011).

Table 3 shows that S.P.'s overall CCMT score was in the 5th percentile, which suggests some recognition deficits with nonface objects. However, a closer examination of her stage-by-stage scores reveals significant learning taking place. S.P.'s scores on Stages 1, 2, and 3 were in the 1st, 5th, and 40th percentiles, respectively. This pattern of improvement indicates that S.P. had initial difficulties with within-class discrimination of nonface objects, but could rapidly overcome these with practice, even for recognition that required generalization across views (Stages 2 and 3 of CCMT). Crucially, this was not the case with her face recognition memory: Across Stages 1, 2, and 3 of the CFMT, S.P. scored in the 1st, 5th, and 4th percentiles, respectively. The clearest contrast between S.P.'s recognition memory of faces and cars can be seen in the comparison of Stages 3 (which uses novel views and lighting, plus additional noise): While S.P.'s Stage 3 score on the CCMT was well within the normal range (40th percentile), her Stage 3 score on the CFMT was not (4th percentile).

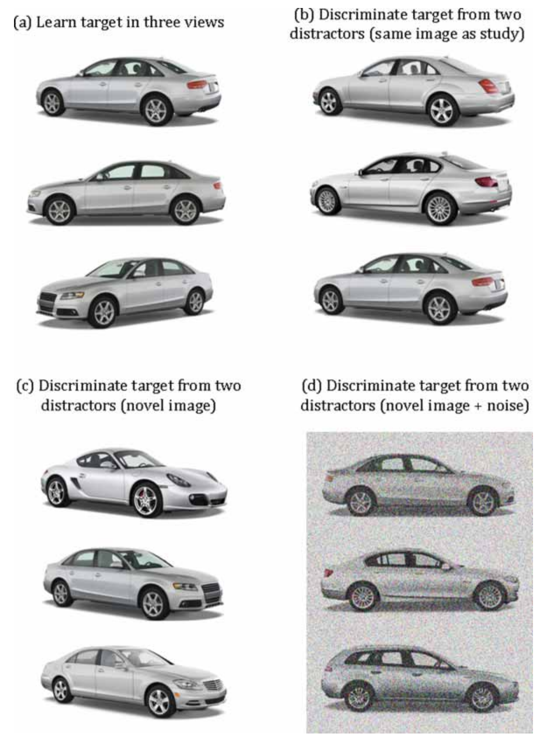

Figure 4. *Illustration of the Cambridge Car Memory Test (using car items not contained in the actual test). To view a colour version of this figure, please see the online issue of the Journal.*

Summary of S.P.'s neuropsychological profile

We conclude that S.P. has developmental prosopagnosia affecting both memory and perception of faces. She may also have some difficulties with within-class recognition of nonface objects, but these problems are not as severe as her deficits in recognizing faces.

EXPERIMENTS 1–3: HOLISTIC PROCESSING

Our aim was to determine whether S.P. is normal or impaired in holistic processing. To overcome the effects of modest task reliability, we used 95% CIs rather than merely point *z* or *t* scores, and we also tested three composite tasks with different sets of face stimuli (see Figure 5) to allow for converging evidence. Experiment 1 used a naming response with three-quarter-view adult faces (McKone, 2008), Experiment 2 used a same/different task with front-view adult face stimuli, and Experiment 3 used a same/different task with front-view child face stimuli (Susilo, Crookes, McKone, & Turner, 2009). To ensure that any composite effect observed in the upright orientation cannot be accounted for by an atypical broad distribution of attention (Figure 1D), we also tested S.P. on Experiments 1 and 2 with the stimuli inverted. All stimulus sets have been previously shown to produce no composite effects for inverted faces in controls.

If S.P. has normal holistic processing, she should show a composite effect for upright faces in all three experiments of similar magnitude to that for controls and no composite effect for inverted faces. However, if her holistic processing is impaired, she should show either (a) an impaired composite effect for upright faces, or (b) a normal composite effect for upright faces in conjunction with a false effect in the same direction for inverted faces.

Method

Experiment 1: Naming task with three-quarter-view adult face stimuli

Our first composite task (taken from McKone, 2008, Experiment 2) used a speeded naming design (Young et al., 1987), adapted for use with lab-based learned faces (Carey & Diamond, 1994). Briefly, six top-half male Caucasian target faces were combined with bottom halves of 10 other individuals to create 60 composites, each shown aligned and unaligned (see Figure 5A). Stimuli were 4° vertical by 3° horizontal (aligned), and 4° by 4.6° (misaligned) at 85 cm viewing distance. In the training phase, participants learned names of the six top-half target faces, with feedback designed to produce ceiling accuracy and high confidence. (Note that S.P. did not require additional training compared to controls: Her naming after four training blocks was above 90%, which was not different from that of controls). The test phase presented the 120 composite stimulus trials (60 aligned, 60 misaligned) in random order, until response. Position on the screen was jittered slightly from trial to trial. Participants were instructed to name the top half as accurately and quickly as possible into a microphone, with reaction time as the measure. Composite scores were calculated as RT aligned minus RT misaligned, excluding trials with RT <250 ms or >3 standard deviations above the subjects' condition mean, or where the name was incorrect. S.P. was tested on the upright version first.

Control data for inverted versions were from McKone (2008, Experiment 2). Controls for upright versions were a new larger sample (81 Caucasians, 18–32 years, 50 females) from the Australian National University community. To calculate split-half reliability of the composite scores, each subject's scores in the aligned condition were split into random halves, the same was done for the misaligned scores, and a composite effect score for each half was calculated. Correlation between the two test halves across participants was calculated and was adjusted using Spearman–Brown correction for list length. This procedure was repeated 50 times, and resulting correlations were averaged to give task reliability of .49 for the upright orientation. Reliability was not calculated for the inverted orientation because controls did not produce a composite effect for inverted faces (i.e., with a mean effect of 0 in the population, individual controls' scores will be merely random variations around 0 even taking the full list of items, before any split-half procedure).

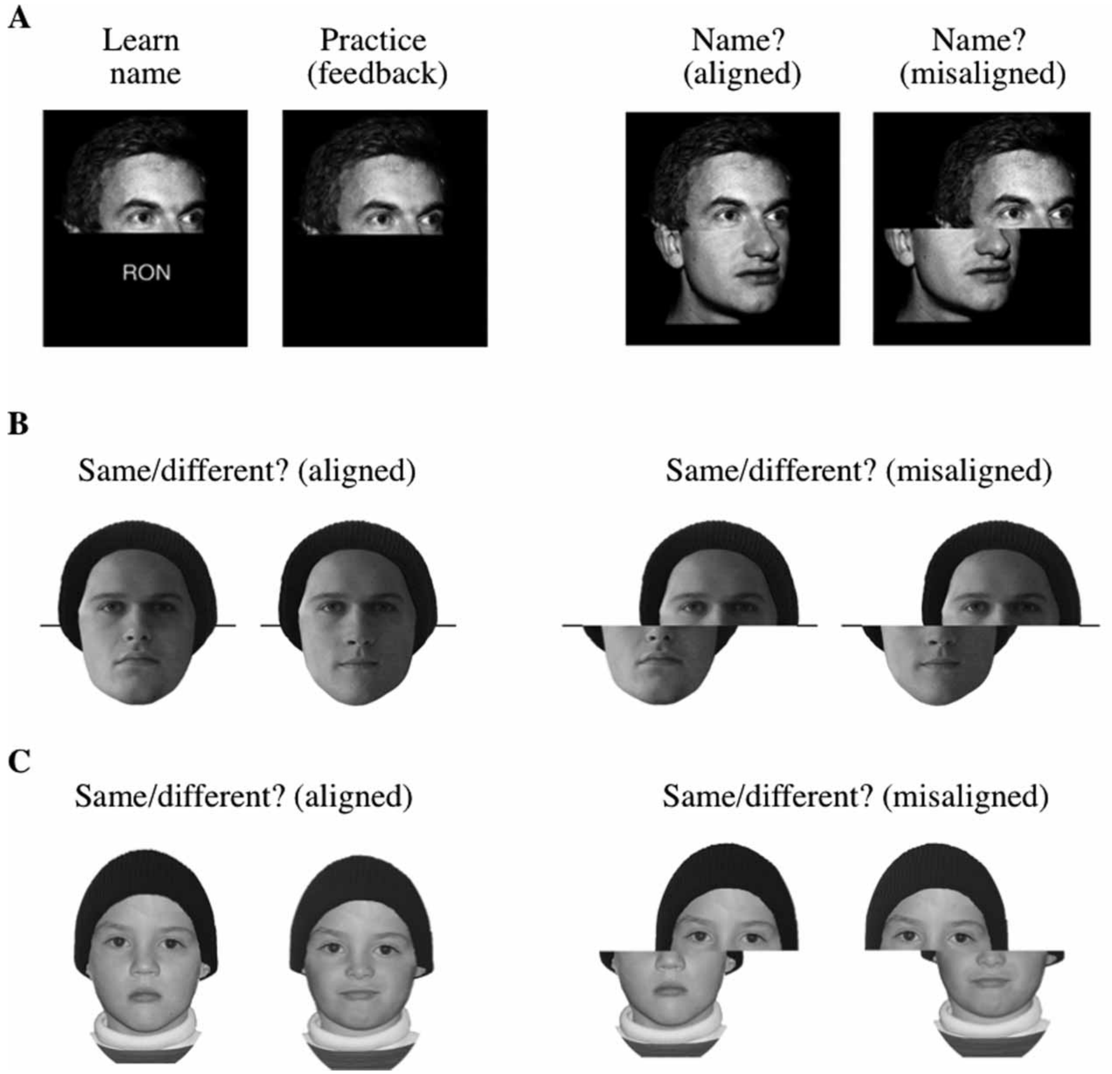

Figure 5. *Example composite faces in (A) Experiment 1, (B) Experiment 2, (C) Experiment 3. In all cases, the task was to respond to the top half of the faces, ignoring the bottom half.*

Experiment 2: Same/different composite task with front-view adult face stimuli

Experiment 2 used a same–different composite task, allowing holistic processing to be measured for completely novel faces. In this version of the procedure (e.g., Le Grand, Mondloch, Maurer, & Brent, 2004), each trial presents two composites (Figure 5B) sequentially. Subjects respond as to whether the top (forehead) half was the *same* (physically identical; half the trials) or *different* (from a different person), ignoring the bottom half. Bottom halves are always different.

We measured the composite effect in the standard way: accuracy for *same–misaligned* trials minus accuracy for *same–aligned* trials (e.g., Le Grand et al., 2004; Michel, Rossion, Han, Chung, & Caldara, 2006; Robbins & McKone, 2007). This approach uses the logic that holistic processing should make it more difficult to see that the top half is the same when aligned with two different bottom halves, and that misalignment breaks this perceptual integration. Note that different trials are not used in the composite score calculation because holistic processing does not make a clear prediction as to whether aligned trials should be more or less accurate than misaligned (direction of the prediction will depend, for example, on the similarity of the two bottom halves relative to the similarity of the two top halves; see Robbins & McKone, 2007). The composite effect on same trials in the same–different task has been directly shown via event-related potentials (ERPs) to reflect a perceptual illusion and not decisional bias (i.e., it is apparent on the perceptual N170 component, and not on the later decisional components P3b and lateralized readiness potential; Kuefner, Jacques, Prieto, Rossion, 2010).

There were 60 original real faces (32 females), all Caucasian front-view greyscale with neutral expressions (Solina, Peer, Bategelj, Juvan, & Kovac, 2003) and Australian National University (ANU) face databases. To facilitate good-quality joins, original faces were first grouped in sets containing individuals with similar colouring and skin tone. Lines at face edge demarcated the halves. A black ski-cap was pasted on to remove hair. Each sequential pair was made in an aligned and a misaligned version. Stimuli were 4.03° vertical by 3.04° horizontal (aligned), and 4.60° by 4.03° (misaligned) at viewing distance 80 cm, against a white background. Stimulus position was jittered (randomly up left, up right, down left or down right, 5° eccentricity from screen centre).

Identical faces were used in upright and inverted blocks. Each orientation had 120 trials (30 each of same–aligned, same–misaligned, different–aligned, and different–misaligned), in random order. Each trial comprised: first composite face for 300 ms; blank screen for 400 ms; second composite face for 300 ms; question "Were the forehead halves same or different?" until response. As the experiment was designed to measure accuracy, no instruction to respond quickly was given. The next trial started after 400 ms. Testing time was 15 min per orientation. Control participants were tested on both upright and inverted orientations, in counterbalanced order (order had no effect on the size of the composite effect for upright faces). S.P. was tested on upright first and then inverted in the same session. Control participants were tested as part of a longer session including bottom-half target trials and other stimulus classes.

Control participants for both upright and inverted conditions were 24 Caucasians (18–33 years, 17 females) from the Australian National University community. Split-half reliability was calculated as described for Experiment 1 (except there were no missing trials, due to measuring accuracy rather than RT). Reliability for upright was .62. Reliability for inverted was not calculated (for same reason as in Experiment 1).

Experiment 3: Same/different composite task with front-view child stimuli

This experiment was as in Susilo et al. (2009). Briefly, stimuli were front-view, neutral-expression, greyscale photographs of Caucasian male children (6–7 years) with black ski-cap (see Figure 5C). Stimulus creation was similar to that in Experiment 2. There were 90 trials (30 same–aligned, 30 same–unaligned, 15 different–aligned, 15 different–unaligned). Participants indicated whether top halves were "same" (physically identical) or "different" (a different person); bottom halves were always different. Viewing distance was 40 cm; aligned faces subtended 9.7° vertical by 6.3° horizontal, and misaligned faces 9.7° × 8.6°; screen position was jittered. Trial procedure was: first composite face for 500 ms; blank screen for 400 ms; second composite face for 500 ms; question "Were the two top-halves same or different?" until response. There was no instruction to respond quickly. Testing time was 15 min.

Both controls and S.P. were tested only on an upright version of the task. Controls were 28 adults in Susilo et al. (2009; aged 22–65 years and averaged here because Susilo et al. found no age effects on the composite scores). Split-half reliability of the composite scores (measured as accuracy for same–misaligned minus accuracy for same−aligned) was calculated as in Experiment 2 and was .75.

Results

S.P.'s performance across Experiments 1−3 is shown in Table 4 and Figure 6. Although, as indicated earlier, there are no clear predictions regarding different trials scores in Experiments 2 and 3, these are reported for completeness in Table 5. Two findings are apparent.

First, for upright faces, S.P. showed normal strength of the composite effects. Figures 6A−6C plot her scores for aligned and misaligned trials separately, compared initially to all controls, and then to a subset of controls who had similar mean performance to that of S.P. in the misaligned condition (the baseline for the composite effect because misaligned scores are not affected by the composite illusion). This latter analysis is important because S.P. tended to name the misaligned face-halves more slowly and match them less accurately than controls, a result not surprising given her prosopagnosia. Table 4 presents S.P.'s percentile ranks relative to both sets of controls. Her scores were always well within the normal range, falling less than one standard deviation from the control means (z scores varied from −0.73 to +0.81 across experiments and control sets). The 95% CIs on S.P.'s percentile rank were, as

Table 4. *Experiments 1−3: S.P.'s composite effect scores and her percentile rank in the population*

				Controls								
Exp.	*Composite task*	*Reliability*	*S.P.*	*M*	*SD*	*N*	z	*Population rank z*	*95% CI of population rank z*	*t*	*Population rank t*	*95% CI of population rank t*
	Upright (all controls)											
1	Naming (ms)	.49	98.39	60.28	93.05	81	0.41	65.91	[15.87, 96.56]	0.41	65.75	[57.20, 73.73]
2	Same/diff adult stimuli (%)	.62	6.67	18.89	16.70	24	−0.73	23.27	[*2.62*, 68.44]	−0.72	24.03	[11.95, 39.22]
3	Same/diff child stimuli (%)	.75	20.00	12.54	13.40	28	0.56	71.23	[33.72, 93.82]	0.55	70.56	[56.10, 82.92]
	Upright (baseline−3matched controls)											
1	Naming (ms)	.49	98.39	105.46	125.11	31	−0.06	47.61	[7.21, 90.99]	−0.06	47.80	[34.15, 61.65]
2	Same/diff adult stimuli (%)	.62	6.67	12.12	12.59	11	−0.43	33.36	[5.05, 78.23]	−0.41	34.36	[14.85, 57.80]
3	Same/diff child stimuli (%)	.75	20.00	8.44	13.30	9	0.87	80.79	[45.62, 96.78]	0.83	78.32	[52.97, 94.81]
	Inverted											
1	Naming (ms)	n/a	−15.63	−21.91	66.66	20	0.09	53.59	n/a	0.09	53.54	[36.88, 70.69]
2	Same/diff adult stimuli (%)	n/a	−30.00	1.49	5.06	24	−6.22	0.00	n/a	−6.10	0.00	[0.00, 0.00]

Note: Composite effect scores, for S.P. and controls, are calculated as disadvantage in the aligned condition compared to the misaligned condition (i.e., *aligned RT − misaligned RT* in Experiment 1 and *misaligned accuracy − aligned accuracy* in Experiments 2 and 3) of same trials (RT = reaction time). Negative z and t scores indicate smaller, or more negative, composite effects than controls. Column labels are as defined in Table 3. Italics indicate values in the lowest 5% of the population.

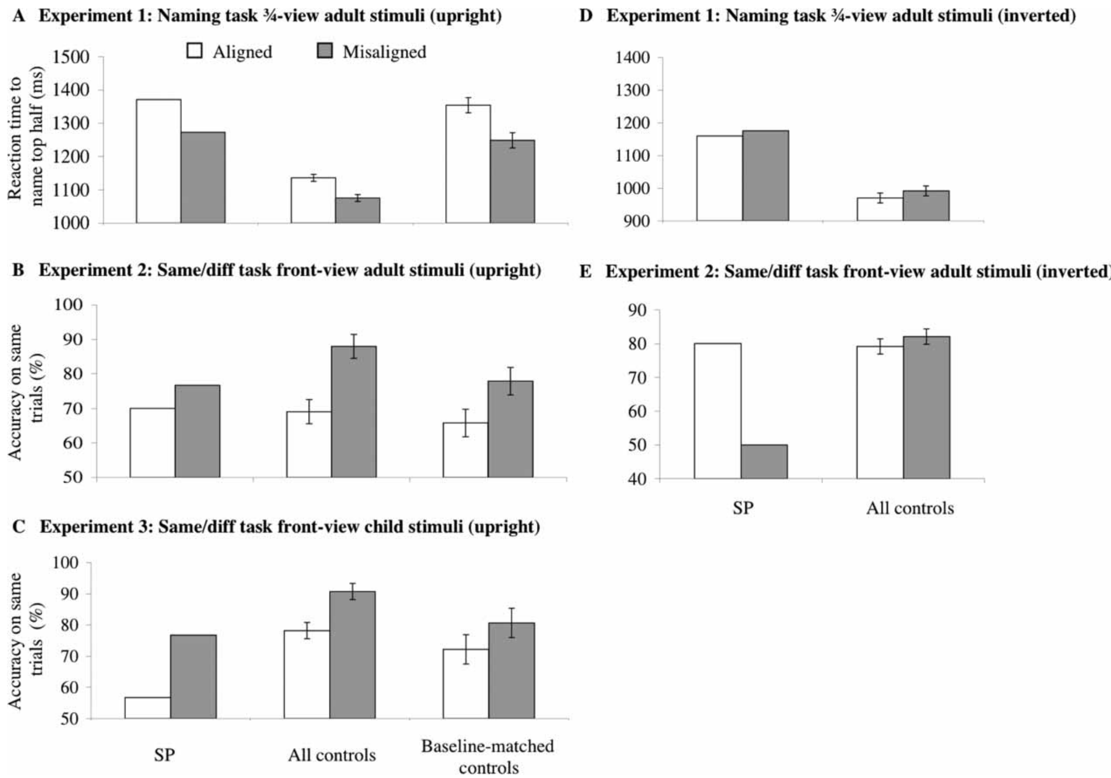

Figure 6. *S.P.'s aligned and misaligned scores in Experiments 1–3, for upright (A–C) and inverted (D–E), relative to all controls, and to baseline-matched controls (i.e., subgroup with mean misaligned equated to S.P.). A positive composite effect is present where performance in the same trials is poorer aligned than misaligned (indicating that ability to perceive the identity of the top-half face is impaired by perceptual integration with a different-identity aligned bottom-half face); that is, aligned bars higher than misaligned on reaction time (plots in A and D), and aligned bars lower than misaligned on accuracy (plots in B, C, and E). Error bars show ± 1 standard error of the difference score between aligned and misaligned, as is relevant for asking whether the control group shows a significant composite effect. Appropriate statistics for asking whether S.P. is different from the control group are given in Table 5 (and show that S.P.'s upright composite scores were in all cases not significantly different from those in either group of controls).*

Table 5. *Different trials scores for S.P. and controls in Experiments 2 and 3*

		S.P. score		*Controls M (SE)*	
Exp.	*Composite task*	*Aligned*	*Misaligned*	*Aligned*	*Misaligned*
2	Same/diff adult stimuli upright (%)	56.67	56.67	75.00 (2.98)	61.81 (3.84)
2	Same/diff adult stimuli inverted (%)	63.33	26.67	63.75 (3.89)	60.69 (3.96)
3	Same/diff child stimuli upright (%)	80	60	85.89 (2.32)	71.15 (4.32)

expected, very broad, particularly those based on task reliability. But even in the worst case scenario (using the lower bound of the reliability-based 95% CI), S.P.'s composite score in Experiments 1 and 3 did not fall in the bottom 5% of the population. Experiment 2 reflected S.P.'s worst performance, yet even here she performed normally compared to baseline-matched controls (z score of −0.43, and lower bound of both 95% CIs not in the bottom 5%).

The second major finding was for inverted faces. These reveal that S.P.'s normal-sized composite effects for upright faces could not be attributed to a false effect arising from unusual distribution of attention. For inverted faces, S.P. showed no evidence of a positive composite effect as would be required to support the attention hypothesis. Instead, in both Experiments 1 and 2 her composite score was, if anything, below zero.

One unexpected finding was that, in Experiment 2, S.P.'s composite score for inverted faces was strongly (and significantly) negative. We have no explanation of why this would be the case. We can think of no theoretical mechanism that could produce a significant negative composite score. Importantly, S.P. showed this abnormal pattern for inverted faces only in Experiment 2: Her pattern in Experiment 1 was perfectly normal (i.e., no difference between aligned and misaligned; see Figure 6D). This argues that S.P. does not have any fundamental abnormality in her processing of inverted faces. Also note that the underlying validity of the method in Experiment 2 is supported by the fact that the controls show the expected pattern of no composite effect inverted (Figure 6E) together with strong positive composite effect upright (Figure 6B).

Discussion

Our findings demonstrate that S.P. had normal levels of holistic processing. This conclusion was based on converging evidence across experiments of a normal-sized composite effect for upright faces, together with a normal lack of a "false" composite effect for inverted faces.

A possibly broader question is whether S.P. is normal at all aspects of "special" face processing. In particular, some authors draw a theoretical distinction between *holistic processing* as tapped by the composite effect and *second-order relational processing* defined as sensitivity to spacing between face features (e.g., Maurer et al., 2002; although note that others have argued against this distinction; McKone & Yovel, 2009; Tanaka & Sengco, 1997). In the present study, we were unable to test S.P. comprehensively on sensitivity to spacing information, but she did show normal sensitivity to changes in one such variable—eye height—in Experiment 5, reported in the next section (particularly see preadaptation discrimination curves in Figure 7A).

EXPERIMENTS 4–6: FACE AFTEREFFECTS IN S.P.

We used a series of aftereffect experiments to assess S.P.'s face space coding. If S.P. has lower sensitivity along one or more dimensions, this would predict aftereffects that are smaller in size than those in controls (see Figure 2). Therefore, if poor face space sensitivity is an origin of S.P.'s prosopagnosia, then she should show smaller face aftereffects than controls. In contrast, if S.P. shows aftereffects as large as controls, then we can infer that she has normal sensitivity of coding in the perceptual space supporting the aftereffects (and also that she paid full attention to the adapting faces, given that attention is required for adaptation; Moradi, Koch, & Shimojo, 2005).

We used three face manipulations. All have been previously used with normal participants and have been shown to employ opponent coding (Dennett et al., 2009; Rhodes & Jeffery, 2006; Robbins et al., 2007). Experiment 4 manipulated identity (e.g., degree of "Dan"-ness, Figure 8A), Experiment 5 eye height (Figure 8B), and Experiment 6 global expansion/contraction (Figure 8C; following Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003).

If S.P.'s face aftereffects are comparable in size to those of controls, this would imply normal coding sensitivity for multiple face attributes. Also, because a size change was included in all experiments, it would indicate that the normal face aftereffects do not arise from low-level vision (although they may not arise solely from face-level representations, an issue we address in Experiments 7–8).

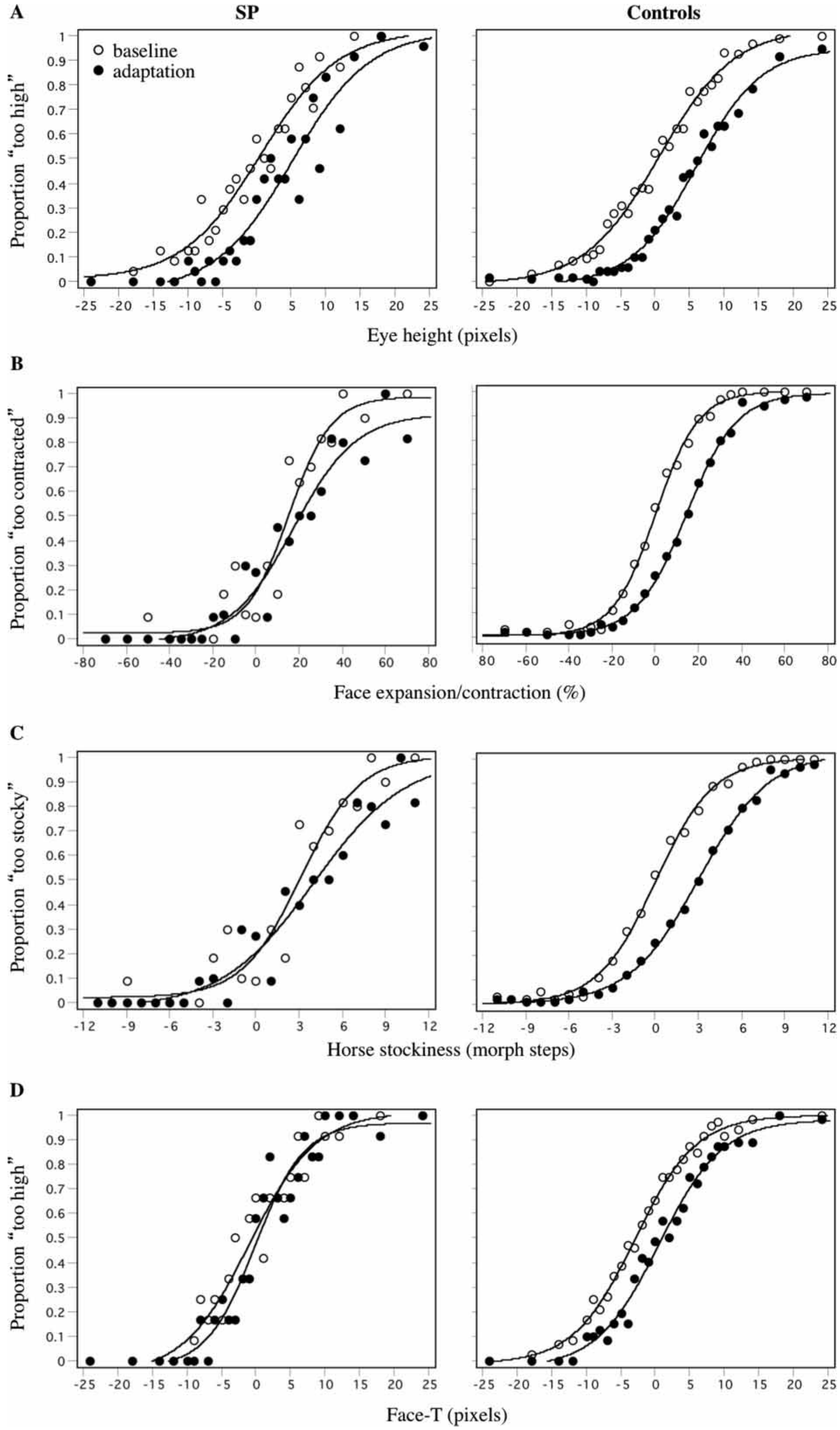

Figure 7. *Psychometric curves for S.P. and the average of controls in (A) Experiment 5, (B) Experiment 6, (C) Experiment 7, and (D) Experiment 8.*

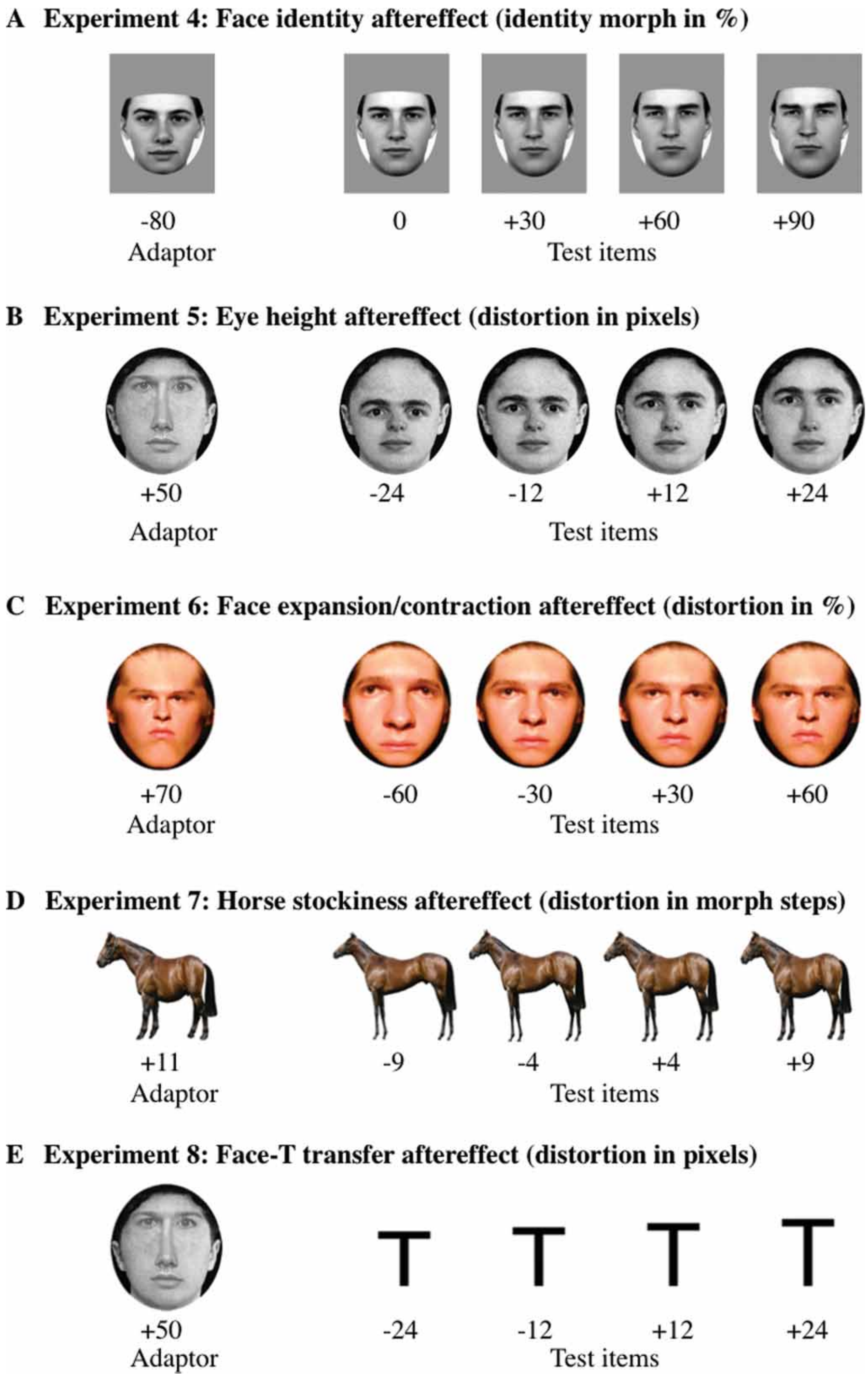

Figure 8. *Stimuli in Experiments 4–8: (A) face identity, (B) eye height, (C) face expansion/contraction, (D) horse stockiness, (E) eye height to T-shape transfer. To view a colour version of this figure, please see the online issue of the Journal.*

Method

Experiment 4: Face identity aftereffect

Subjects categorized faces as either "Team Dan" or "Team Jim", with stimuli being either Dan or Jim (two different individuals) in four identity-strength versions (0, 30, 60, 90% strength of each individual; note 0% is the average face made via morphing together a large number of individuals, see Figure 8A). An aftereffect was indicated by more "Team Dan" responses after adapting to anti-Dan than after adapting to anti-Jim conditions (i.e., adapting to anti-Dan biases participants to see Dan, whereas adapting to anti-Jim biases them to see Jim; Pellicano et al., 2007). The procedure was the same as that in Nishimura et al. (2010) with the exceptions of (a) size change between adapt and test, (b) small differences in test-face identity strength values and trial numbers, and (c) no baseline (unadapted) condition.

A training phase presented 40, 60, and 100% versions of Dan and Jim until participants correctly named 40 and 60% versions in at least 4 out of 5 trials. In the test phase, trial procedure was: adaptor face (80%-strength anti-Dan or 80%-strength anti-Jim) for 5,000 ms; 150 ms interstimulus interval (ISI); test face for 400 ms; grey screen during which participants indicated whether the test face was "Team Dan" or "Team Jim". Thre were 84 trials: 7 (identity strengths; 0%, plus 3 of Dan and 3 of Jim) × 6 (repeats) × 2 (adapt faces). Adaptor stimuli were 7.8° vertical by 7.8° horizontal, test stimuli 6.2° by 5.8° at 50 cm viewing distance. All stimuli were presented at screen centre. Task duration was 30 min. Size of the aftereffect was calculated as difference between the two adapting conditions (*adapt anti-Dan* − *adapt anti-Jim*) in overall proportion of "Team Dan" responses; this measure has been used previously when, as here, there are relatively few identity-strength values, and so it is difficult to fit psychometric curves and determine the face perceived as most normal in individual subject data (Jiang, Blanz, & O'Toole, 2007; Nishimura, Maurer, Jeffery, Pellicano, & Rhodes, 2008; Rhodes, Evangelista, & Jeffery, 2009).

Control participants were 12 Macquarie University community members (7 females; aged 20−59 years, but combined because we found no relationship between age and size of aftereffect, $r = -.03$, $p = .92$). Split-half reliability of the proportion correct scores was calculated for these participants, based on dividing the six trials at each identity strength value randomly in two then computing total proportion of "Dan" responses collapsed across identity strength. The Spearman−Brown corrected split-half reliability was .59.

Experiment 5: Eye height aftereffect

This experiment was the same as that in Susilo, McKone, and Edwards (2010b, Experiment 1, 50-pixel adaptor condition). Briefly, four individual faces were used as adaptors, and a different four as test stimuli, all on a common background head. Adaptor stimuli had eyes moved up to close to hairline (" + 50 pixels"). Test stimuli were eyes moved up (coded as positive) or down (negative) in 29 steps (Figure 8B). Adaptors were 7.9° vertical by 5.7° horizontal, test stimuli 10° by 7.9°, viewing distance 40 cm, all stimuli at screen centre. Baseline (*no-adaptation*) phase trial procedure was: test stimulus for 250 ms; forced-choice question "Were the eyes too high or too low?" (relative to subjects' imagined real-world average) until response; 400 ms delay before next trial. The *with-adaptation* phase was identical except that each test stimulus was preceded by 4,000 ms adaptor stimulus (outlined by black rectangle for easy differentiation), then 400 ms blank screen. Each phase had 348 trials: 29 (eye height levels) × 4 (test faces) × 3 (repeats); total trials for 2 phases = 696.

Size of the aftereffect was determined via pre- versus post-adaptation shift in the point of subjective equality (PSE; the physical stimulus perceived as having the most normal eye height). Each subject's data in each phase were plotted as a psychometric curve (% "too-high" responses as a function of the 29 eye height levels) and were fitted using the logistic function. The PSE was defined as the stimulus eye height corresponding to 50% "too high" responses from the fit. Because adaptation to our eyes-up face (+50 pixels) should move the baseline PSE toward higher eyes stimuli, the size of the aftereffect was calculated as *adapted PSE − baseline PSE*. Thus, as in all subsequent experiments in this article, a positive score indicates a shift in the direction reflecting an aftereffect.

Control subjects were 20 Caucasians (18−31 years, 12 females; 16 from Susilo et al., 2010b, Experiment 1; 4 from Susilo et al., 2010a). Split-half reliability of aftereffect size, using 50 splits and Spearman−Brown correction, was .85. To derive the input data for the split-half correlations, each subject's 12 trials per eye height value (i.e., 3 repeats of 4 test individuals) in the no-adaptation baseline were split in two (i.e., into 2 groups of 6), resulting in two plots from which separate baseline PSEs were calculated; the same was done for the with-adaptation condition to obtain two adapted PSEs, and these were subtracted to obtain two aftereffect shift scores. Controls were tested on one run through the experiment (1 hr duration). S.P. was tested on two runs, in sessions separated by 4 weeks, and these were averaged; this would be expected to give S.P.'s aftereffect score even greater reliability than that of controls.

Experiment 6: Face expansion/contraction aftereffect

We used the global expansion/contraction manipulation of Rhodes et al. (2003). Unlike previous studies using expansion/contraction, only one individual face was used (as both the adaptor and test); the rationale was that when we tested aftereffects for nonface objects in Experiment 7, our manipulation was applied to only one horse, and hence we wanted one of our face aftereffect procedures to be similar. The single face was a front-view neutral expression male (from ANU face database) in an elliptical aperture that included inner hairline. Adobe Photoshop 5.5 "spherize" function created 23 steps (±70, ±60, ±50, ±40, ±35, ±30, ±25, ±20, ±15, ±10, ±5, 0% spherized) from extremely expanded (−70%) to extremely contracted (+70%; Figure 8C). The adaptor was always the extreme +70% contraction. At viewing distance 60 cm the adaptor was 17° vertical by 14° horizontal, test faces 12° by 10° . To further reduce low-level retinotopic contributions to the aftereffects, adaptor faces were at screen centre, and test faces randomly at four locations 3.5° from screen centre (top left, top right, bottom left, bottom right). All stimuli were against a white background.

Baseline phase trial procedure was: test face for 600 ms; forced-choice question "Was the face too expanded or too contracted?" until response via keyboard buttons; 400 ms intertrial interval. In the *adaptation phase*, participants viewed the adaptor face for 2 min (with instructions to move eyes around). *Postadaptation* procedure was identical to baseline, except that each trial had the adaptor face presented for 2,000 ms before the test face. Each phase had 240 trials: ≥10 (repeats) × 23 (stimulus levels), for total of 480 trials. Task duration was 45 min.

Size of adaptation aftereffect was calculated as shift in PSE (*postadaptation PSE − baseline PSE*), using the curve-fitting procedure described for Experiment 5. Control participants were 12 Australian National University community members (18−44 years, 8 females) paid $10 or given first-year psychology course credits. Split-half reliability was calculated as in Experiment 5, and it was .85.

Results and discussion

In the identity aftereffect task, both controls and S.P. reached ceiling accuracy for the most extreme values (i.e., close to 0% "Dan" responses for the 90%-Jim face and close to 100% "Dan" responses for the 90%-Dan face), with a smooth change in responses in between. The same is apparent for the eye height and expansion/contraction tasks (Figures 7A and 7B). In the latter two tasks, the measure of the aftereffect was PSE shift in psychometric curves, and so we first confirmed that S.P.'s data produced good-quality fits. For eye height, S.P.'s fits gave R^2 of .91 without adaptation, and .95 with adaptation; for expansion/contraction, S.P.'s fits gave R^2 of .94 without adaptation, and .94 with adaptation (all these were in the normal range of controls, mean R^2 = .93, range = .89−.99).

Table 6. *Experiments 4–8: S.P.'s aftereffect results and her percentile rank in the population*

Exp.	*Aftereffect task*	*Reliability*	*S.P.*	*Controls M*	*Controls SD*	*Controls N*	*z*	*Population rank z*	*95% CI of population rank z*	*t*	*Population rank t*	*95% CI of population rank t*
	Faces											
4	Face identity (proportion correct)	.59	0.17	0.11	0.09	12	0.67	74.86	[27.76, 97.26]	0.64	73.25	[51.01, 90.04]
5	Eye height (pixels)	.85	4.54	6.27	2.88	20	−0.60	27.43	[8.69, 56.36]	−0.59	28.23	[14.19, 45.36]
6	Face expansion/contraction (%)	.85	6.62	9.66	10.49	12	−0.29	38.59	[14.69, 68.08]	−0.28	39.29	[19.44, 61.59]
	Object											
7	Horse stocky/thin (morph steps)	.81	2.92	2.69	2.16	22	0.11	54.38	[22.66, 83.15]	0.10	54.10	[37.68, 70.00]
	Transfer											
8	Eye height to T-shape transfer (pixels)	.62	0.92	1.79	1.92	6	−0.45	32.64	[*4.95*, 77.64]	−0.42	36.59	[5.22, 65.16]

Note: The reported measure is the aftereffect size (e.g., computed as the shift in the point of subjective equality—the face/horse/T perceived as most normal—following adaptation; see Method section for details). Direction of coding is that negative *z* and *t* scores indicate aftereffect size smaller than controls. "Population rank *z*" and "population rank *t*" refer to the percentage of control population who perform worse than S.P. as calculated using *z* scores (i.e., ignoring control sample size) and *t* scores (i.e., taking into account control sample *N*). "95% CI on population rank *z*" uses internal task reliability to give the 95% confidence interval on S.P.'s percentile rank. "95% CI on population rank *t*" uses control sample *N* to give the 95% confidence interval on S.P.'s percentile rank. Italics indicate values in the lowest 5% of the population (note that in the one case where this occurred, the direction indicates a *larger* F−T, face to T, transfer aftereffect in S.P., i.e., if anything, slightly more face-specificity than controls).

Table 6 demonstrates that S.P.'s aftereffects for face stimuli were normal in size compared to controls. On all three tasks, her face aftereffects were within 0.6 standard deviations of the control mean. She showed a slightly larger aftereffect than the control mean on one task (identity, 63rd percentile) and slightly smaller on the other two tasks (eye height, 28th percentile; expansion/contraction 39th). The lower bound of S.P.'s 95% CI did not include impaired performance using either task reliability or control sample size to determine the confidence interval, and even taking poorest 5% as a lenient definition of impairment.

We conclude from Experiments 4–6 that S.P. had normal sensitivity of coding along dimensions within some type of "perceptual space" that is responsive to face stimuli. Given the size change between adapt and test (plus position change in Experiment 6), S.P.'s aftereffects did not arise from low-level vision. However, at this stage it remains open whether her aftereffects arose specifically from face space, or from shape-generic midlevel vision processes (Regan & Hamstra, 1992; Suzuki, 2005).

EXPERIMENTS 7–8: DO S.P.'S FACE AFTEREFFECTS ARISE FROM FACE SPACE REPRESENTATIONS?

Experiments 7 and 8 were designed to address the extent to which S.P.'s normal-sized aftereffects for face stimuli (Experiments 4–6) arose specifically from a high-level face space as opposed to shape-generic components present in midlevel vision. In typical controls, aftereffects for upright face stimuli have been shown to derive from a combination of these two mechanisms (Susilo et al., 2010a; for broader discussion, see Rhodes & Leopold, 2011). Theoretically, the contribution of shape-generic mechanisms is to be expected because manipulations made to face stimuli not only alter face information per se, but also alter basic shape components that are known to show aftereffects in isolation. For example, contracting the width of the mouth alters vertical-to-horizontal aspect ratio, a basic midlevel vision shape property known to demonstrate aftereffects (Regan & Hamstra, 1992).

In two experiments, we assessed whether S.P.'s normal-sized face aftereffects could have derived from an atypically large shape-generic contribution to her aftereffects. That is, we tested the hypothesis that S.P. had a *stronger* shape-generic component, and a *weaker* face-specific component, than in controls, which add to give the same-sized total aftereffect for face stimuli.

In Experiment 7, we examined S.P.'s aftereffects for a nonface object class (horses, see Figure 8D). Our question was whether S.P. showed an aftereffect for horses that was significantly larger than normal. If this were the case, it would provide preliminary evidence consistent with the view that S.P. has larger-than-normal shape-generic aftereffects.

In Experiment 8, we provided a more direct test of the relative proportions of face-specific and shape-generic contributions to S.P.'s face aftereffects. We did so by testing the amount of transfer of the aftereffect from a face adaptor to a nonface test stimulus. We used the eye-height manipulation, because this has a corresponding basic shape manipulation that fully captures the type of midlevel information altered in the face—the height of a T stimulus matched to the T-shaped face region that comprises eyes (horizontal bar) and nose–mouth (vertical line). In controls, we have previously shown that approximately 45% of the eye height aftereffect transfers to a *t* test stimulus; that is, *adapt-face* then *test-T* (F–T) is 45% as large as *adapt-face* then *test-face* (F–F; Susilo et al., 2010a). This demonstrates that in controls, the total face aftereffect arises approximately half from face-specific mechanisms (i.e., *face* space) and half from neural mechanisms shared between faces and Ts (i.e., *shape* space). If S.P.'s face aftereffect arises disproportionately from shape-generic mechanisms, her aftereffect in the F–T condition should be larger than that in controls, and her "transfer score" (F–T as a proportion of F–F) should be larger than the 45% shown by controls.

Method

Experiment 7: Horse stockiness aftereffect

We selected horses as stimuli because they share important properties with faces, particularly being natural objects that have curves rather than straight lines and smooth rather than abrupt boundaries between parts. We also chose a manipulation type (i.e., a "stocky" versus "thin" distortion) that is similar to the expanded–contracted face manipulation (Experiment 6) in that it affects all regions of the stimulus simultaneously, and it produces similar types of distortion at the local level (e.g., regions becoming more or less vertically elongated).

The experiment (Dennett et al., 2009) was identical in procedure to Experiment 6. We used images of a thoroughbred horse and a Shetland pony taken from the internet, morphed together and caricatured using Gryphon Morph 2.5, to generate 23 equally-spaced morph levels ranging from extremely thin to extremely stocky (Figure 8D). The adaptor was the most stocky image. The adaptor was 17° vertical by 20° horizontal, the test stimuli 14° by 17°; viewing distance was 60 cm. The task was to decide whether the horse stimulus was "too thin" or "too stocky". Task duration was 45 min.

Control participants were 22 Australian National University community members (18–31 years, 15 females). Split-half reliability was calculated as in Experiment 5, and it was .81.

Experiment 8: Eye height to T-shape transfer

This experiment was identical to Experiment 5, except that we now assessed how much of the face adaptation transferred to a T stimulus. The adaptor stimulus was a face with eyes shifted up by 50 pixels, but the test stimuli (pre- and postadaptation) were Ts rather than faces. The experiment is described in detail in Susilo et al. (2010a, Experiment 2; using procedure of Susilo et al., 2010b, Experiment 1, 50-pixel adaptor condition). Briefly, T-shape stimuli (Figure 8E) were created from an Arial font capital "T", matched in size to the internal eyes–nose–mouth "T" region of the face stimuli (i.e., such that the zero-value T overlaid the eyes–nose–mouth region of the zero-value faces); a 1-pixel change in height of the bar in the T was physically identical to a 1-pixel change in eye height. The task was to indicate whether the T was too tall/short compared to their imagined average T. Task duration was 1 hr.

Control participants were 6 Caucasians from the Australian National University community (20–31 years, 3 females; from Susilo et al., 2010a, Experiment 2). Aftereffect score was calculated as in Experiment 5, and it was .62.

Results

Figure 7 shows S.P.'s psychometric curves pre- and post-adaptation for horse (Figure 7C) and face-to-T (Figure 7D) experiments. S.P.'s R^2 values were .93 preadaptation and .96 postadaptation for horses, and .95 preadaptation and .96 postadaptation for F–T transfer, all of which are comparable to controls (across both experiments, control mean = .94, range .87–.98).

Table 6 shows the results. Findings did not support larger than normal shape-generic aftereffects in S.P. For horses, S.P.'s aftereffect was only .11 standard deviations larger than the mean. For face-to-T, the tendency was for S.P.'s aftereffect to be smaller than that of the control mean (although still within half a standard deviation), indicating, if anything, a slightly greater face-specific proportion of the face eye height aftereffect in S.P. than in controls (i.e., opposite to the direction predicted by a larger than normal shape-generic origin). Overall, the size of S.P.'s aftereffects on both tasks was completely normal, with no suggestion of either abnormally large or abnormally small effects based on the upper and lower bounds of the 95% CIs.

We also computed S.P.'s transfer score (i.e., F–T as a proportion of her aftereffect in the corresponding F–F condition from Experiment 5) and did the same for the 6 controls who did both experiments. Again, S.P.'s trend was towards, if anything, weaker transfer across shapes than most controls: a transfer proportion of .19 for S.P. and mean of .59 for controls.

Discussion

Taken together, Experiments 4–8 have shown that S.P.'s face aftereffects were comparable in size and nature to those of controls: Her face aftereffects could not be accounted for solely by either low-level mechanisms or shape-generic processes. We conclude that S.P.'s face aftereffects, like those of controls, reflect normal face space coding. This in turn argues that S.P. has a face space in which the response functions of the opponent model are of normal steepness, which implies normal sensitivity to changes in values along each tested manipulation. Because we tested three different manipulations (identity, eye height, expansion/contraction) that presumably project onto multiple face space dimensions (whatever these might be), it seems reasonable to conclude that her normal coding sensitivity applies broadly across face space.

One issue is whether S.P. might show deficits in face space coding if she were tested using tasks other than face aftereffects. We see no reason to suspect she would. Our results argue that S.P. possesses the coding sensitivity required to produce at least two of the standard face space effects—namely, the caricature effect and average-is-attractive effect. Specifically, S.P.'s face aftereffects demonstrate she has a normal representation of both the average face (thereby predicting an average-is-attractive effect) and also of exactly how far any exemplar face is from this average along a given trajectory (thereby implying normal coding of how much a face has been caricatured). Further, although we cannot completely rule out that S.P.'s face space coding might differ in some subtle way from that of controls,[3] we think this is unlikely. S.P.'s results demonstrate that her face space (a) discriminates faces as accurately as controls when the stimuli fall *along a trajectory that runs through the average value*, such as in the eye height and expanded/contracted tasks, and (b) also has normal representation of the relative location of *faces that fall on different trajectories*. This latter idea is evident in the identity task, in which adapting to anti-Dan made S.P. see more Dan in Jim stimuli (as well as in neutral or weak Dan faces) to the same extent as did controls. This implies that the relative locations of Jim and Dan were normally represented.

GENERAL DISCUSSION

Our results demonstrate that S.P. has severe face recognition deficits, present on both face memory and face perception tests. These deficits could not be attributed to developmental failure of either holistic processing or face space. S.P. showed normal composite effects for upright and not for inverted faces in three tasks (Experiments 1–3), and normal coding of "spacing" information for eye–nose distance (Experiment 5); she also showed normal-sized face aftereffects for manipulations of identity, eye height, and expansion/contraction that were generated by normal face-level mechanisms rather than low-level or shape-generic processes (Experiments 4–8). The case of S.P. shows that it is possible for normal holistic processing and face space coding to exist despite a lifetime of impaired face recognition.

What mechanisms could potentially be impaired in S.P.?

Given that S.P.'s holistic processing and face space coding were no different than those in controls, there must be some other mechanism/s in which she is impaired in order to account for her prosopagnosia. Here we suggest two possibilities: problems with part-based processing, and/or with view/image generalization. Note that these suggestions are speculative; we were unable to conduct further empirical tests with S.P. at the present time.

One possibility is impaired *part-based processing*. A possible impairment in S.P.'s part-based processing for general objects is perhaps suggested by her performance on the Cambridge Car Memory Test (which was impaired on Stage 1), although this would need to be confirmed by further direct tests. Future studies could also test for impairment specifically in processing of face parts (e.g., tell apart a sequence of noses in isolation, Nunn, Postma, & Pearson, 2001; or memory in the part-alone condition of the part–whole task, Tanaka & Farah, 1993).

The second possibility is impairment in *generalization across view and/or other image changes*. All of the tests on which S.P. demonstrates impaired performance (i.e., the diagnostic tests for prosopagnosia) require the ability to generalize across face images, by ordering front-view faces relative to three-quarter-view ones (CFPT), or remembering learned targets in images containing novel views and/or lighting (CFMT), or recognizing famous faces in novel poses (MACCS FFT). In contrast, all tasks on which we found S.P. to be normal did not require view/image generalization. Each composite task tested only one given view (either front or three quarters). Similarly, our face aftereffect tasks used only one view (front). It is possible that, although S.P. is normal on holistic processing and face space in single views, she is somehow impaired in the use of one or both mechanisms across views. We are not aware of a

[3]For example, S.P. has not been tested as to whether she shows identity-contingent aftereffects (e.g., adaptation to contraction for one identity and simultaneously expansion for another; Robbins & Heck, 2009; Yamashita, Hardy, DeValois, & Webster, 2005).

method of testing the role of holistic processing in view generalization. However, it may be possible to test the role of face space coding in view generalization, by examining the view tuning of face aftereffects (i.e., transfer of aftereffects from one view to another). In normal participants, expansion/contraction aftereffects are reduced by about half with a 45° difference between adapt and test viewpoints (Jeffery, Rhodes, & Busey, 2006), and face identity aftereffects transfer more strongly across view change for familiar than for unfamiliar faces (Jiang et al., 2007). Potentially, S.P. might show narrower tuning than controls, or lack the view-breadth advantage for familiar compared to unfamiliar faces.

If S.P. does have a view generalization problem, it is currently unclear whether this problem would be limited to faces. The fact that S.P. performed normally on Stage 3 of the Cambridge Car Memory Test (which requires generalizing cars across views) suggests that her problem may be face-specific. However, cars tend to have more abrupt part boundaries than faces. It remains possible that S.P. might show impaired view generalization of objects that are more similar to faces in their visual structure (e.g., horses, which, like faces, have gradual rather than abrupt borders between parts). Also, in everyday life, S.P. reports problems suggestive of difficulty in representing space in three dimensions (e.g., finding it hard to learn to drive; see the Case Description).

The importance of task reliability in single-case studies of prosopagnosia

An important secondary finding of our study is that the reliability of tasks similar to those used in previous DP studies is only modest: The reliability of the three composite tasks was in the range of .49 to .75, and for the identity aftereffect task it was .59. Only two tasks (not previously used with DP) had good reliability: our eye height aftereffect task (.85) and our face expansion/contraction aftereffect task (.85).

We therefore suggest the following approaches for future single-case studies. First, researchers should regularly report internal task reliability and 95% CIs around point z scores and use these to inform single-case inferences. For example, a suitable criterion (as used here) would be that performance is considered normal only if the lower bound of the case's 95% CI does not fall in the impaired range. Second, whenever possible, researchers should attempt to modify tasks developed for group-based experiments so that they have good reliability for single-case inferences. This can sometimes be achieved by increasing the number of trials. For example, in the present study, there is nothing to suggest that the eye height aftereffect is inherently more reliable than the identity aftereffect. It is likely that the former was more reliable than the latter because of the number of trials (692 vs. 84). Where increasing number of trials is not feasible, one alternative is to use multiple versions of a task (with different stimulus sets) measuring the same construct, allowing for converging evidence.

Implications of S.P. for other cases of DP

Given the evidence that DP is a heterogeneous disorder, it is important to note that there is no reason to think that because S.P. shows one pattern—intact holistic processing and intact face space—all other DPs should also have the same pattern. Indeed, for holistic processing, there is already clear previous evidence of heterogeneity. Two recent group studies have reported that, on average, the DP group showed significantly less holistic processing than controls ($N = 14$, Avidan et al., 2011; $N = 12$, Palermo, Willis, et al., 2011), indicating that at least some individual DPs must have impaired holistic processing.[4] However, other DPs appear to be like S.P. in showing normal levels of holistic

[4]And also indicating that, in contrast to the suggestion of Konar et al. (2010), holistic processing can be functionally associated with face individuation ability.

processing: Le Grand et al. (2006) reported 7 out of 8 of the DPs they tested showed normal composite effects using point *z* scores, and this conclusion was supported in our own analysis (Table 1) for 4 of these individuals even using the lower bound *z* (which we estimated presuming reliability of .65). Similarly, for face space, although the one previous publication reported normal-sized aftereffects in DP (Nishimura et al., 2010), a recent conference presentation reported a small but significant reduction in mean identity aftereffect in a group analysis ($N = 14$; Palermo, Rivolta, Wilson, & Jeffery, 2011), again implying that at least some DP individuals are impaired.

Overall, we suspect that S.P. may turn out to be quite a rare case in showing both holistic processing and face space mechanisms to be intact. Her case is important, however, in that recent research has focussed strongly on holistic processing and (single-view) face space coding as the "mechanisms to test" in DP. The existence of a case who can be completely normal in both these mechanisms yet very impaired in face recognition argues that studies of DP would benefit from investigating a broader range of mechanisms potentially relevant to face recognition. These include part-based processing and view/image generalization; we gave some suggestions for how to test these mechanisms at the beginning of this section.

CONCLUSION

S.P. is a case of prosopagnosia whose theoretical importance is to provide an "existence proof". Her case demonstrates, for the first time, that it is possible for face recognition ability to be very poor despite perfectly normal abilities in the two perceptual components that have been investigated most commonly in previous studies of developmental prosopagnosia and are most widely presumed to be functional contributors to face recognition ability in typically developing participants.

S.P.'s case demonstrates that normal holistic processing and face space are not, even when present together, sufficient to ensure good face recognition (although note that this does not mean that they are not necessary). In S.P.'s case, there must be other important component/s of face processing that should have developed but failed to do so. More broadly, it is possible that the components missing in S.P. might also be missing and contribute to prosopagnosia in other cases of DP. We therefore argue that the field should consider adding other candidate mechanisms to the battery investigated as possible causes of developmental prosopagnosia.

Manuscript received 16 February 2011
Revised manuscript received 21 July 2011
Revised manuscript accepted 25 July 2011
First published online 10 November 2011

REFERENCES

Avidan, G., Tanzer, M., & Behrmann, M. (2011). Impaired holistic processing in congenital prosopagnosia. *Neuropsychologia*, *49*, 2541−2552.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5−17.

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130–1149.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, *58,* 382–398.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., et al. (2009). Diagnosing prosopagnosia: Effects of aging, sex, and participant−stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455.

Busigny, T., Joubert, S., Felician, O., Ceccaldi, M., & Rossion, B. (2010). Holistic perception of the individual face is specific and necessary: Evidence from an extensive case study of acquired prosopagnosia. *Neuropsychologia*, *48*, 4057–4092.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition*, *1,* 253–274.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40,* 1196–1208.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *The Clinical Neuropsychologist*, *23,* 1173–1195.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482–486.

Dennett, H., Edwards, M., & McKone, E. (2009). Are objects like faces? Norm-based versus exemplar-based coding as revealed by adaptation aftereffects. *Journal of Vision*, *9*(8), 518a.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2011). *The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects.* Manuscript submitted for publication.

Duchaine, B. (2011). Developmental prosopagnosia. In A. J. Calder, G. Rhodes, M. H. Johnston, & J. V. Haxby (Eds.), *Handbook of face perception*. New York, NY: Oxford University Press.

Duchaine, B., & Nakayama, K. (2006a). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic patients. *Neuropsychologia*, *44,* 576–585.

Duchaine, B., & Nakayama, K. (2006b). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16,* 166–173.

Duchaine, B., Yovel, G., & Nakayama, K. (2007). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social Cognitive and Affective Neuroscience*, *2,* 104–113.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*, 1187–1196.

Gryphon Software Corporation. (1993). *Morph's user guide.* San Diego: Gryphon Software Corporation.

Herzmann, G., Danthir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*(3), 840–857.

Jeffery, L., Rhodes, G., & Busey, T. (2006). View-specific coding of face shape. *Psychological Science*, *17*(6), 501–505.

Jiang, F., Blanz, V., & O'Toole, A. J. (2007). The role of familiarity in three-dimensional view-transferability of face identity adaptation. *Vision Research*, *47,* 525–531.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science*, *21,* 38–43.

Kuefner, D., Jacques, C., Prieto, E. A., & Rossion, B. (2010). Electrophysiological correlates of the composite face illusion: disentangling perceptual and decisional components of holistic face processing in the human brain. *Brain and Cognition*, *74*, 225–238.

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., et al. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, *61,* 139–158.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science*, *15*(11), 762–768.

Ley, P. (1972). *Quantitative aspects of psychological assessment: An introduction*. New York, NY: Harper & Row.

Maddess, T., McCourt, M. E., Blakeslee, B., & Cunningham, R. B. (1988). Factors governing the adaptation of cells in area-17 of the cat visual cortex. *Biological Cybernetics*, *59,* 229–236.

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

McKone, E. (2008). Configural processing and face viewpoint. *Journal of Experimental Psychology: Human Perception and Performance*, *34,* 310–327.

McKone, E., Aimola Davies, A., Fernando, D., Aalders, R., Leung, H., Wickramariyaratne, T., et al. (2010). Asia has the global advantage: Race and visual attention. *Vision Research*, *50*(16), 1540–1549.

McKone, E., & Robbins, R. (2011). Are faces special? In A. J. Calder, G. Rhodes, M. H. Johnston, & J. V.

Haxby (Eds.), *Handbook of face perception*. Oxford, UK: Oxford University Press.

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Towards a new theory of holistic processing. *Psychonomic Bulletin & Review*, *16*(5), 778–797.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of our own race. *Psychological Science*, *17,* 608–615.

Moradi, F., Koch, C., & Shimojo, S. (2005). Face adaptation depends on seeing the face. *Neuron*, *45,* 169–175.

Nishimura, M., Doyle, J., Humphreys, K., & Behrmann, M. (2010). Probing the face-space of individuals with prosopagnosia. *Neuropsychologia*, *48,* 1828–1841.

Nishimura, M., Maurer, D., Jeffery, L., Pellicano, E., & Rhodes, G. (2008). Fitting the child's mind to the world: Adaptive norm-based coding of facial identity in 8-year-olds. *Developmental Science*, *11*(4), 620–627.

Nunn, J. A., Postma, P., & Pearson, R. (2001). Developmental prosopagnosia: Should it be taken at face value. *Neurocase*, *7,* 15–27.

Palermo, R., Rivolta, D., Wilson, C. E., & Jeffery, L. (2011, May). *Abnormal adaptive coding of identity in congenital prosopagnosia*, Paper presented at Vision Sciences Society meeting, Naples, FL.

Palermo, R., Willis, M. L., Rivolta, D., McKone, E., Wilson, C. E., & Calder, A. J. (2011). Impaired holistic coding of facial expression and facial identity in congenital prosopagnosia. *Neuropsychologia*, *49,* 1226–1235.

Pellicano, E., Jeffery, L., Burr, D., & Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Current Biology*, *17,* 1–5.

Ramon, M., Busigny, T., & Rossion, B. (2010). Impaired holistic processing of unfamiliar individual faces in acquired prosopagnosia. *Neuropsychologia*, *48,* 933–944.

Raven, J. C., Raven, J., & Court, J. H. (1991). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford, UK: Oxford Psychologists Press.

Regan, D., & Hamstra, S. J. (1992). Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research*, *32*(10), 1845–1864.

Rhodes, G., Evangelista, E., & Jeffery, L. (2009). Orientation-sensitivity of face identity aftereffects. *Vision Research*, *49,* 2379–2385.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46,* 2977–2987.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: Face adaptation and attractiveness aftereffects. *Psychological Science*, *14*(6), 558–566.

Rhodes, G., & Leopold, D. A. (2011). Adaptive norm-based coding of face identity. In A. J. Calder, G. Rhodes, M. H. Johnston, & J. V. Haxby (Eds.), *Handbook of face perception*. Oxford, UK: Oxford University Press.

Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham Object Recognition Battery*. Hove, UK: Lawrence Erlbaum Associates.

Robbins, R., & Heck, P. (2009). Brad Pitt & Jude Law: Individual-contingent face aftereffects and norm- versus exemplar-based models of face-space. *Journal of Vision*, *9*(8), 516a.

Robbins, R., & McKone, E. (2003). Can holistic processing be learned for inverted faces? *Cognition*, *88,* 79–107.

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, *103*(1), 34–79.

Robbins, R., McKone, E., & Edwards, M. (2007). Aftereffects for face attributes with different natural variability: Adaptor position effects and neural models. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 570–592.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica*, *123*(2), 274–289.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257.

Schmalzl, L., Palermo, R., & Coltheart, M. (2008). Cognitive heterogeneity in genetically based prosopagnosia: A family study. *Journal of Neuropsychology*, *2,* 99–117.

Solina, F., Peer, P., Bategelj, B., Juvan, S., & Kovac, J. (2003). *Color-based face detection in the "15 seconds of fame" art installation*. Paper presented at the Conference on Computer Vision/Computer Graphics Collaboration for Model-based Imaging, Rendering, Image Analysis and Graphical Special Effects, March 2003, Roquencourt, France.

Susilo, T., Crookes, K., McKone, E., & Turner, H. (2009). The composite task reveals stronger holistic processing in children than adults for child faces. *PLoS ONE*, *4*(7), e6460.

Susilo, T., McKone, E., & Edwards, M. (2010a). Solving the upside-down puzzle: Why do upright and inverted face aftereffects look alike? *Journal of Vision*, *10*(13), 1.

Susilo, T., McKone, E., & Edwards, M. (2010b). What shape are the neural response functions underlying opponent coding in face space? A psychophysical investigation. *Vision Research*, *50,* 300–314.

Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and after-effects in high-level vision* (pp. 135–172). Oxford, UK: Oxford University Press.

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

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25,* 583–592.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology A*, *43*(2), 161–204.

Van Belle, G., de Graef, P., Verfaillie, K., Rossion, B., & Lefèvre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision*, *10*(5), 10.

Webster, M. A., & MacLeod, D. I. A. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B*, *366,* 1702–1725.

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin and Review*, *6*(4), 647–653.

Wickramariyaratne, T. (2006). *The other-race effect in face recognition: Contact, holistic processing, and field dependence*, Unpublished honours thesis, Australian National University, Canberra, Australia.

Williams, M. A., Berberovic, N., & Mattingley, J. B. (2007). Abnormal fMRI adaptation to unfamiliar faces in a case of developmental prosopamnesia. *Current Biology*, *17,* 1259–1264.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is highly heritable. *Proceedings of the National Academy of Sciences*, *107,* 5238–5241.

Wilson, C. E., Palermo, R., Rivolta, D., Williams, M., & Schmalzl, L. (2008). *Macquarie Centre for Cognitive Science Famous Face Test 2008.* Unpublished manuscript.

Yamashita, J. A., Hardy, J. L., DeValois, K. K., & Webster, M. A. (2005). Stimulus selectivity of figural aftereffects for faces. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 420–437.

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

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., et al. (2010). Heritability of the specific cognitive ability of face perception. *Current Biology*, *20,* 1–6.