

SCIENTIFIC REPORTS



OPEN

Atypical Time Course of Object Recognition in Autism Spectrum Disorder

Laurent Caplette¹, Bruno Wicker² & Frédéric Gosselin¹

Received: 17 May 2016
Accepted: 30 September 2016
Published: 18 October 2016

In neurotypical observers, it is widely believed that the visual system samples the world in a coarse-to-fine fashion. Past studies on Autism Spectrum Disorder (ASD) have identified atypical responses to fine visual information but did not investigate the time course of the sampling of information at different levels of granularity (i.e. Spatial Frequencies, SF). Here, we examined this question during an object recognition task in ASD and neurotypical observers using a novel experimental paradigm. Our results confirm and characterize with unprecedented precision a coarse-to-fine sampling of SF information in neurotypical observers. In ASD observers, we discovered a different pattern of SF sampling across time: in the first 80 ms, high SFs lead ASD observers to a higher accuracy than neurotypical observers, and these SFs are sampled differently across time in the two subject groups. Our results might be related to the absence of a mandatory precedence of global information, and to top-down processing abnormalities in ASD.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder whose most prominent symptoms are deficits in social interaction and communication, restricted interests and repetitive behaviors. However, ASD is also characterized by sensory and perceptual peculiarities, as research has increasingly demonstrated in recent years¹. In particular, several studies have reported that ASD subjects are biased toward visual local information compared to global information^{2–4}. According to some authors, this is the result of a deficit in the extraction of global aspects of images, which is argued to be a defining feature of autism⁵. Other authors propose that there is no such deficit and that ASD subjects are only better at accessing local information⁶. A recent meta-analysis of these few studies found no strict deficit in global processing or enhanced local processing, but rather only a delay in the processing of global information, suggesting that it is the temporal aspect of this process that is different in ASD⁷. Atypical spatial frequency (SF) processing is also observed in ASD. High SFs are related to local information as they are both particularly well suited to represent small objects in a scene; and low SFs are related to global information as they are both particularly well suited to represent large objects in a scene (they are not equivalent, however, because global information can be retrieved from high SFs). Unlike neurotypicals, individuals with ASD exhibit a greater P1 amplitude in response to high SFs than to low SFs⁸, and a reduced difference between neural responses to high and intermediate SFs^{9,10}. It was also shown that individuals with ASD use high SFs more than low SFs to recognize faces, while neurotypicals do the opposite². Other studies reported that the contrast sensitivity function (CSF; the low-level visual sensitivity to sinusoidal gratings at different SFs) of ASD subjects peaks at higher SFs than that of neurotypicals^{11,12}, albeit this remains controversial^{13,14}.

Importantly, SFs are not sampled all at once during recognition in neurotypicals: they are extracted in a coarse-to-fine manner^{15,16}. Rapidly extracted low SFs may furthermore guide, in a top-down manner, the subsequent extraction of higher SFs^{17–19}. Interestingly, there is evidence that individuals with ASD rely less on top-down mechanisms during perception, when compared to neurotypical individuals, and that this might be crucial in explaining the phenotype of autism^{20–23}. If individuals with ASD do not rely on top-down mechanisms, they might not need to sample SFs in a coarse-to-fine fashion.

To date, however, the time course of SF sampling has never been investigated in ASD. Moreover, most studies demonstrating a coarse-to-fine sampling of SF information in neurotypicals compared a single low SF condition to a single high SF condition, which is unhelpful to uncover the *exact* SFs at play in this process. Here, we used a novel technique based on the Bubbles approach²⁴ to map with unprecedented precision the use of SFs across

¹Département de psychologie, Université de Montréal, Montreal, Canada. ²Laboratoire de Neurosciences Cognitives, CNRS UMR 7291, Aix-Marseille Université, Fédération 3C, Marseille, France. Correspondence and requests for materials should be addressed to L.C. (email: laurent.caplette@umontreal.ca)

time in both neurotypical and ASD subjects. On each trial, subjects had to recognize an object from a brief video sampling random SFs on random frames; we then reverse correlated the revealed SFs with response accuracy. This technique allows us to reveal the difference between the groups for each SF in each time frame, and thus to settle major issues such as whether ASD subjects exhibit a deficit in sampling low SFs, an enhanced sampling of high SFs, or a difference in the time course of SF sampling.

Results

There were no difference between groups in average response time (ASD: 814 ms; Controls: 752 ms; $t(58) = 0.77$, $p > 0.25$) or in the average quantity of information (the number of bubbles; see Methods) needed to recognize objects with 75% accuracy (ASD: 42.29; Controls: 40.06; $t(58) = 0.39$, $p > 0.25$).

To uncover which spatial frequencies in which time frames led to accurate object recognition, we performed multiple least-square linear regressions between accuracies and the SF x time sampling matrices (which indicate if each SF in each time frame is sampled or not) of the corresponding trials, for each subject. Then, for each SF on each time frame, we compared the regression coefficients across subjects and computed one-sample and two-sample t statistics; this resulted in within-group and between-group t maps indicating respectively how the presentation of specific SFs on specific time frames correlates with accuracy for each group and how these correlations differ between the groups (see 2D plots on Fig. 1).

In neurotypical subjects, SFs between about 1 and 20 cycles per image (cpi) throughout stimulus presentation and SFs up to 36 cpi in the second half of the video ($p < 0.05$, corrected for Family-Wise Error Rate (FWER); peak Cohen's $d = 1.60$) led to accurate recognition. In ASD subjects, three blobs reached statistical significance ($p < 0.05$, FWER-corrected; peak $d = 1.90$): the largest (291 pixels) peaked at 11 cpi and 80 ms, the second largest (132 pixels) at 15.5 cpi and 296 ms, and the smallest (63 pixels) at 3.5 cpi and 213 ms. In the group contrast, SFs peaking at 25.5 cpi and 71 ms led to more accurate recognition in ASD participants than in neurotypical subjects ($p < 0.05$, FWER-corrected; peak $d = 1.16$; Fig. 1).

To investigate the time course of sampling of each SF and uncover if it was used increasingly or decreasingly across time, we fitted a line across time for each SF in each subject map. We then took the slopes of these lines across subjects and computed one-sample t statistics for each SF. This resulted in within-group t vectors indicating the slope of use (increasing or decreasing) of each SF across time for each group (Fig. 1, upper panel, 1D plots). To test whether the time course of sampling was different between the groups, we then computed two-sample t statistics for each SF. This resulted in a between-group t vector indicating how the slopes of use differed between the groups (Fig. 1, lower panel, 1D plots).

Results show that while ASD subjects seem to use each SF in an approximately constant manner across time, neurotypical subjects use SFs around 25 cpi increasingly across time ($p < 0.05$, FWER-corrected; peak $d = 0.57$). These same SFs are used more in an increasing manner across time in neurotypical subjects than in ASD subjects ($p < 0.05$, FWER-corrected; peak $d = 1.08$; Fig. 1). This confirms an interaction between group and time course of sampling of SFs around 25 cpi and shows that this time course differed between the groups for these SFs.

Discussion

In this study, we investigated with unprecedented precision how neurotypical and ASD subjects sample SFs across time. We first observed that neurotypicals use lower SFs (up to about 20 cpi) throughout all stimulus presentation, while they use higher SFs increasingly with the passage of time. This confirms previous findings of a coarse-to-fine sampling, and nuances them by showing that low SFs continue to be used in the latest time frames. This is the first piece of evidence for a hypothesis originally formulated by Ullman²⁵ (see also ref. 26) according to which low SFs are continuously extracted to activate coarse representations of new objects in the visual field.

Moreover, we have discovered that ASD observers do not extract SFs across time like neurotypical observers: in the first 80 ms, high SFs around 25 cpi lead ASD observers to a higher accuracy than neurotypical observers, and these SFs are sampled differently across time in the two subject groups. An increased early sampling of high SFs in ASD subjects might explain the enhanced neural activity they exhibit early in visual processing in response to such SFs⁸. Furthermore, their initial simultaneous sampling of low and high SFs is consistent with the observation of a reduced segregation between neural responses to intermediate and high SFs in their early processing^{9,10}.

Some authors have suggested that autistic perception could be explained by a deficit in the processing of global information⁵; others suggested that it could be explained better by a superiority in accessing local information⁶. Published findings on the subject are contradictory^{1,7,27}. Importantly, our results do not show any difference in the extraction of low SFs (analogous to global information), and only a more immediate extraction of high SFs (analogous to local information). These results suggest that the difference in SF perception between neurotypicals and ASD observers relies specifically in the time course of extraction of high SFs. The local bias present in ASD perception might thus be the result of a more immediate access to fine visual information without the mandatory precedence of global information observed in neurotypicals.

The absence of a coarse-to-fine sampling in individuals with ASD might also be linked to abnormalities in top-down processing^{20,22}. Prominent models of object recognition state that the extraction of low SFs activates an object representation that modulates the subsequent extraction of higher SFs in a top-down fashion¹⁷⁻¹⁹. In ASD observers, early extraction of low SFs cannot guide the sampling of the diagnostic higher SFs since high and low SFs are sampled equally early, soon after stimulus onset.

Methods

Participants. Forty-three neurotypical adults and twenty-one adults with ASD were recruited. Power analyses were not performed before the recruitment to determine the optimal sample sizes because of the use of a new experimental technique and unknown effect sizes. A sample size of approximately 20 was chosen for the ASD group because it is fairly common in studies of this clinical population; a larger sample size of approximately 40

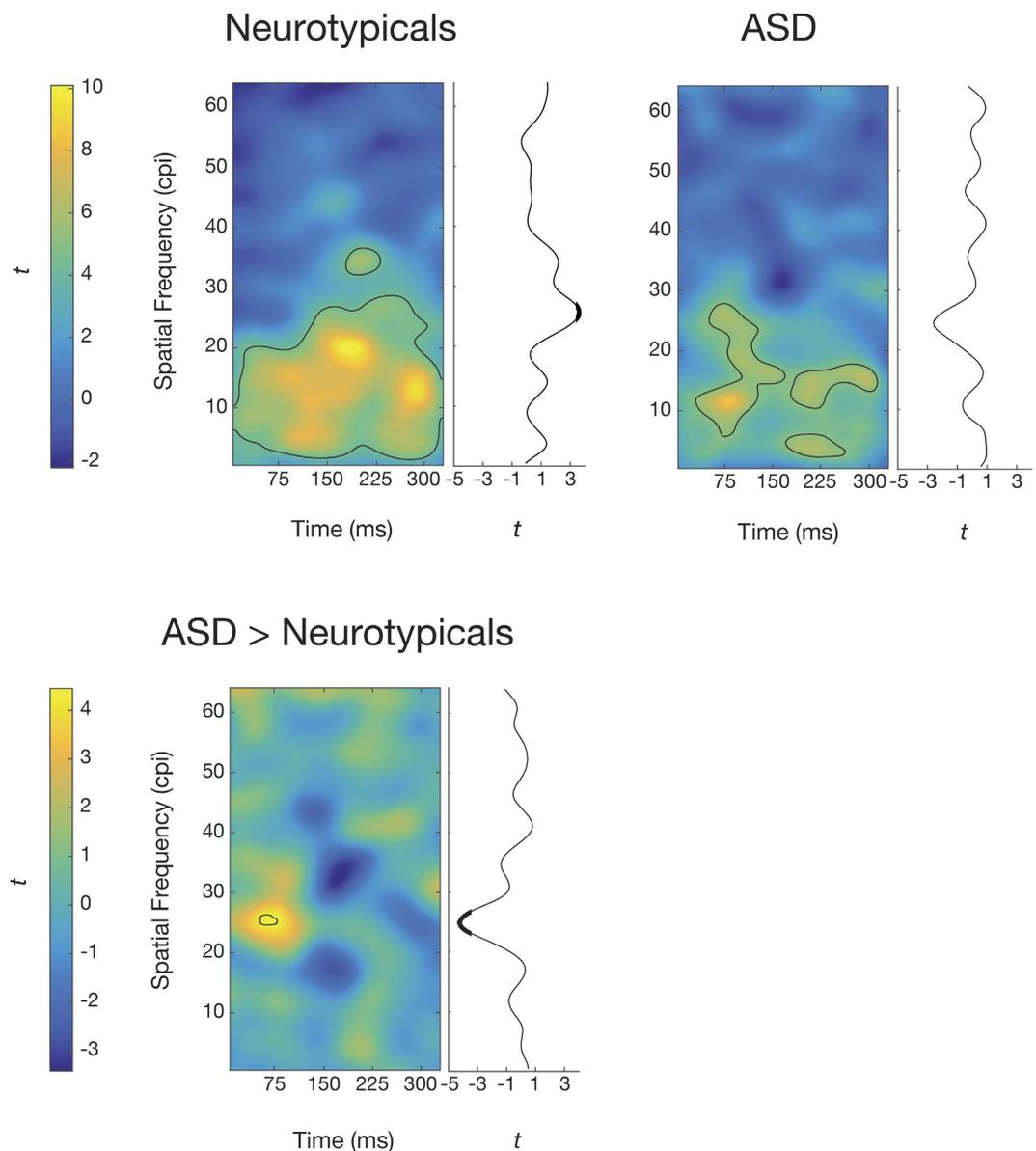


Figure 1. Use of SFs across time in each group and differences between the groups. Upper panel: One-sample t maps illustrating how each SF in each time frame correlates with accurate object recognition, and one-sample t vectors indicating the slope of use of each SF across time, for the neurotypical and ASD groups. Lower panel: Two-sample t map illustrating the between-group differences in the use of each SF in each time frame, and two-sample t vector indicating the between-group differences in the slope of use of each SF across time. Pixels enclosed by black lines and bold portions of the vectors are significant ($p < 0.05$, FWER-corrected). Note that since the width of all images subtended 6 degrees of visual angle, cycles per image can be converted in cycles per degree by dividing by 6. Note also that only statistics up to 64 cycles per image (cpi) are shown; there are no statistically significant results in the portions not shown. Note finally that color axes are different in the panels.

was chosen for the neurotypical group to increase our statistical power. ASD participants were diagnosed by an expert psychiatrist or licensed clinical psychologist and the diagnosis had to be recently confirmed, with each having met the criteria for ASD within the past 3 years on the basis of the DSM-IV-TR²⁸. In addition, most participants were also tested using either the AQ, ASAS-R or ADI-R questionnaire. Neurotypical participants were recruited on the campus of the Université de Montréal as a comparison group. Brief interviews ensured that none of the participants suffered from any mental or neurological disorder (other than ASD for the ASD group) and that they were free of medication.

Three neurotypical participants were excluded prior to the analysis: one because he did not complete the first block, one because the quantity of information he required to reach target performance was more than 3 standard deviations over the group mean, and one because his mean response time was more than 3 standard deviations over the group mean. One ASD participant was excluded because the quantity of information he required to reach target performance was more than 3 standard deviations over the group mean. The final ASD group thus included

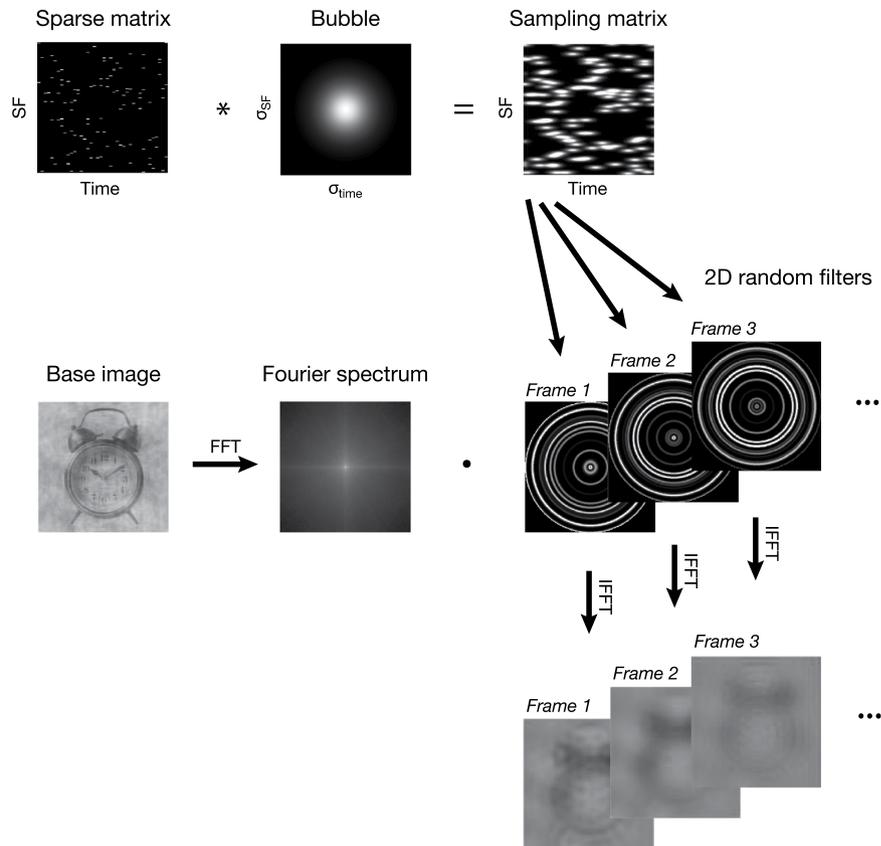


Figure 2. Illustration of the sampling method. On each trial, we randomly generated a matrix of dimensions 256×40 (representing respectively SFs and frames) in which most elements were zeros and a few were ones. We then convolved this sparse matrix with a 2D Gaussian kernel (a “bubble”). This resulted in the trial’s sampling matrix, shown here as a plane with a number of randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic 2D random filters. Finally, these 2D random filters were dot-multiplied by the base image’s spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame.

20 participants (14 males; mean age = 27.10, SD = 8.72), and the final neurotypical group included 40 participants (19 males; mean age = 24.43, SD = 7.55). Subject groups did not differ significantly in age ($t(58) = 1.17, p > 0.25$) or gender ($p = 0.17$, Fisher’s exact test). Participants of both groups had or were completing a post-secondary diploma at the time of the study, and had normal or corrected-to-normal vision. The experimental protocol was approved by the ethics board of the University of Montreal’s Faculty of Arts and Sciences and the study was carried in accordance with the approved guidelines. Written informed consent from all participants was obtained after the procedure had been fully explained, and a monetary compensation was provided upon completion of the experiment.

Materials. The experimental program ran on Mac Pro (Apple Inc.) computers in the Matlab (Mathworks Inc.) environment, using functions from the Psychophysics Toolbox^{29,30}. All stimuli were presented on Asus VG278H monitors (1920×1080 pixels at 120 Hz), calibrated to allow linear manipulation of luminance. Luminance values ranged from 1.6 cd/m^2 to 159 cd/m^2 . Chin rests were used to maintain viewing distance at 76 cm.

Stimuli. Eighty-six grayscale object images were selected from the database used in Shenhav, Barrett, & Bar³¹ and from Internet searches. Images were 256×256 pixels (6×6 degrees of visual angle) and median object width was 220 pixels (SD = 47 pixels). The objects were cropped manually and pasted on a uniform mid-gray background. The SF spectrum of every object image was set to the average SF spectrum across all object images using the SHINE toolbox³². This preserved the most important spectral properties of natural objects while eliminating sources of undesired variance in the stimuli. Resulting images had a root mean square (RMS) contrast of about 0.20.

On each trial, participants were shown a short video (333 ms) consisting of an object image with random SFs gradually revealed at random time points (e.g., Video S1; Video S2). To create these stimuli, we first randomly generated, on each trial, a matrix of dimensions 256×40 (representing respectively SFs from 0.5 to 128 cycles per image or cpi, and frames, each lasting 8.33 ms) in which most elements were zeros and a few were ones. The number of ones was adjusted on a trial-by-trial basis to maintain performance at 75% correct. We then convolved this sparse matrix with a 2D Gaussian kernel (a “bubble”; $\sigma_{\text{SF}} = 1.5 \text{ cpi}$; $\sigma_{\text{time}} = 15 \text{ ms}$). This resulted in the

trial's sampling matrix: a SF \times time plane with randomly located bubbles. Every column of this sampling matrix was then rotated around its origin to create isotropic 2D random filters. Finally, these 2D random filters were dot-multiplied by the base image's spectrum and inverse fast Fourier transformed to create a filtered version of the image for every video frame (see Fig. 2 for an illustration of this method). Thus, a SF \times time bubble centered on the SF of 10 cpi and the frame 6 in the sampling matrix would reveal this SF on this frame (and, to a lesser extent, the neighboring SFs on the neighboring frames) *in the entire image*. To ensure accurate luminance display, we applied noisy-bit dithering to the final stimuli³³.

This random sampling method was preferred to other methods with lower dimensionality because it allowed us to perform a systematic search of the SF \times time space and thus to precisely infer which SFs in which time frames led to an accurate response. Moreover, this method is unbiased, as it does not require the selection of a small number of arbitrary cut-offs.

Procedure. After they had completed a short questionnaire for general information (age, sex, lateralisation, education, language), participants sat in front of a computer monitor, in a dim-lighted room. Participants completed two 500-trial blocks on the first day and two more on another day. Each trial comprised the following consecutive events, on a mid-gray background: a fixation cross (300 ms), a uniform mid-gray field (200 ms), the video stimulus (333 ms), a fixation cross (300 ms), a uniform mid-gray field (200 ms), and an object name that remained on the screen until a response was provided by the participant or for a maximum of 1 s. Subjects were asked to indicate whether the name matched the object (it did 50% of the time) with a keyboard key press as rapidly and accurately as possible. The number of bubbles (i.e. the quantity of information revealed during the whole video) was adjusted on a trial-by-trial basis using a gradient descent algorithm to maintain performance at 75% correct.

Regression analysis. Before performing the regressions, accuracies and response times were transformed in z-scores for each object (to minimize variability due to psycholinguistic factors), for each 500-trial block (to minimize variability due to task learning), and for each subject (to minimize residual individual differences in performance). Trials associated with z-scores over 3 or below -3 (either in accuracy or response time) were discarded (2.13% of trials overall). Sampling matrices were also transformed in z-scores on each trial; this equalized the importance given to all trials, irrespective of the number of bubbles used on these trials.

Following the regressions, the subject-wise maps of regression coefficients were convolved with a Gaussian kernel ($\sigma_{\text{SF}} = 3.5$ cpi; $\sigma_{\text{time}} = 42$ ms) and transformed in z-scores with a bootstrapped sample, before computation of the *t* statistics. The two-sample *t* statistics were computed according to Welch's formula, which does not assume equal variances of the populations. Statistical significance of the *t* maps was assessed with a Pixel test³⁴, which controls the FWER while taking into account the correlation in the data. All reported statistical test results are two-tailed.

References

1. Simmons, D. R. *et al.* Vision in autism spectrum disorders. *Vision Res.* **49**, 2705–2739 (2009).
2. Deruelle, C., Rondan, C., Gepner, B. & Tardif, C. Spatial frequency and face processing in children with autism and Asperger syndrome. *J. Autism Dev. Disord.* **34**, 199–210 (2004).
3. Plaisted, K., Swettenham, J. & Rees, L. Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *J. Child Psychol. Psych.* **40**, 733–742 (1999).
4. Rinehart, N. J., Bradshaw, J. L., Moss, S. A., Brereton, A. V. & Tonge, B. J. Atypical interference of local detail on global processing in high-functioning autism and asperger's disorder. *J. Child Psychol. Psych.* **41**, 769–778 (2000).
5. Happé, F. & Frith, U. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* **36**, 5–25 (2006).
6. Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J. Autism Dev. Disord.* **36**, 27–43 (2006).
7. Van der Hallen, R., Evers, K., Brewaeys, K., Van den Noortgate, W. & Wagemans, J. Global processing takes time: A meta-analysis on local–global visual processing in ASD. *Psychol. Bull.* **141**, 549–573 (2015).
8. Vlamings, P. H. J. M., Jonkman, L. M., van Daalen, E., van der Gaag, R. J. & Kemner, C. Basic abnormalities in visual processing affect face processing at an early age in autism spectrum disorder. *Biol. Psych.* **68**, 1107–1113 (2010).
9. Boeschoten, M. A., Kenemans, J. L., Engeland, H. V. & Kemner, C. Abnormal spatial frequency processing in high-functioning children with pervasive developmental disorder (PDD). *Clin. Neurophysiol.* **118**, 2076–2088 (2007).
10. Jemel, B., Mimeault, D., Saint-Amour, D., Hoesin, A. & Mottron, L. VEP contrast sensitivity responses reveal reduced functional segregation of mid and high filters of visual channels in autism. *J. Vision* **10**, 13 (2010).
11. Kéïta, L., Guy, J., Berthiaume, C., Mottron, L. & Bertone, A. An early origin for detailed perception in autism spectrum disorder: biased sensitivity for high-spatial frequency information. *Sci. Rep.* **4**, 5475 (2014).
12. de Jonge, M. V. *et al.* Visual information processing in high-functioning individuals with autism spectrum disorders and their parents. *Neuropsychology*, **21**, 65–73 (2007).
13. Koh, H. C., Milne, E. & Dobkins, K. Spatial contrast sensitivity in adolescents with autism spectrum disorders. *J. Autism Dev. Disord.* **40**, 978–987 (2010).
14. Behrmann, M. *et al.* Configural processing and its relationship to face processing. *Neuropsychologia* **44**, 110–129 (2006).
15. Hughes, H. C., Nozawa, G. & Kitterle, F. Global precedence, spatial frequency channels, and the statistics of natural images. *J. Cogn. Neurosci.* **8**, 197–230 (1996).
16. Schyns, P. G. & Oliva, A. From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* **5**, 195–200 (1994).
17. Bar, M. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* **15**, 600–609 (2003).
18. Bar, M. *et al.* Top-down facilitation of visual recognition. *Proc. Natl Acad. Sci. USA* **103**, 449–454 (2006).
19. Goddard, E., Carlson, T. A., Dermody, N. & Woolgar, A. Representational dynamics of object recognition: Feedforward and feedback information flows. *Neuroimage*, **128**, 385–397 (2016).
20. Loth, E., Gómez, J. C. & Happé, F. When seeing depends on knowing: Adults with autism spectrum conditions show diminished top-down processes in the visual perception of degraded faces but not degraded objects. *Neuropsychologia* **48**, 1227–1236 (2010).
21. Ropar, D. & Mitchell, P. Shape constancy in autism: the role of prior knowledge and perspective cues. *J. Child Psychol. Psych.* **43**, 647–653 (2002).

22. Pellicano, E. & Burr, D. When the world becomes “too real”: a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* **16**, 503–509 (2012).
23. Van de Cruys, S. *et al.* Precise minds in uncertain worlds: Predictive coding in autism. *Psychol. Rev.* **121**, 649–675 (2014).
24. Gosselin, F. & Schyns, P. G. Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res.* **41**, 2261–2271 (2001).
25. Ullman, S. Visual routines. *Cognition* **18**, 97–159 (1984).
26. Caplette, L., McCabe, E., Blais, C. & Gosselin, F. The time course of object, scene and face categorization in *Handbook of categorization in cognitive science, 2nd edition* (eds C. Lefebvre & H. Cohen) (Elsevier, in press).
27. Behrmann, M. *et al.* Configural processing in autism and its relationship to face processing. *Neuropsychologia* **44**, 110–129 (2006).
28. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th edition, text revision (Author, 2000).
29. Brainard, D. H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433–436 (1997).
30. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437–442 (1997).
31. Shenhav, A., Barrett, L. F. & Bar, M. Affective value and associative processing share a cortical substrate. *Cogn. Affect. Behav. Neurosci.* **13**, 46–59 (2013).
32. Willenbockel, V. *et al.* Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Meth.* **42**, 671–684 (2010).
33. Allard, R. & Faubert, J. The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution. *Behav. Res. Meth.* **40**, 735–743 (2008).
34. Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M. & Gosselin, F. Accurate statistical tests for smooth classification images. *J. Vision* **5**, 659–667 (2005).

Acknowledgements

This study was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN/04777-2014) to F.G., and by scholarships from the Social Sciences and Humanities Research Council of Canada (SSHRC) and from the Fonds de recherche du Québec Nature et technologies (FRQNT) to L.C. We are grateful to Philippe Des Roches for his help with data collection. We thank all ASD volunteers and their families for participating in the project.

Author Contributions

F.G., B.W. and L.C. designed the experiment. L.C. analyzed the data, prepared the figures, and wrote the main manuscript text. All the authors (F.G., B.W. and L.C.) reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Caplette, L. *et al.* Atypical Time Course of Object Recognition in Autism Spectrum Disorder. *Sci. Rep.* **6**, 35494; doi: 10.1038/srep35494 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016