HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2015 November 27.

Published in final edited form as:

Neuron. 2015 November 4; 88(3): 604–616. doi:10.1016/j.neuron.2015.09.042.

Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking

Shuo Wang[1,2,*], **Ming Jiang**[3,*], **Xavier Morin Duchesne**[4], **Elizabeth A. Laugeson**[5], **Daniel P. Kennedy**[4], **Ralph Adolphs**[1,2], and **Qi Zhao**[3]

[1]Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA

[2]Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA

[3]Department of Electrical and Computer Engineering, National University of Singapore, 117583 Singapore

[4]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

[5]Department of Psychiatry and PEERS Clinic, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90024, USA

Summary

The social difficulties that are a hallmark of autism spectrum disorder (ASD) are thought to arise, at least in part, from atypical attention towards stimuli and their features. To investigate this hypothesis comprehensively, we characterized 700 complex natural scene images with a novel 3-layered saliency model that incorporated pixel-level (e.g., contrast), object-level (e.g., shape), and semantic-level attributes (e.g., faces) on 5551 annotated objects. Compared to matched controls, people with ASD had a stronger image center bias regardless of object distribution, reduced saliency for faces and for locations indicated by social gaze, yet a general increase in pixel-level saliency at the expense of semantic-level saliency. These results were further corroborated by direct analysis of fixation characteristics and investigation of feature interactions. Our results for the first time quantify atypical visual attention in ASD across multiple levels and categories of objects.

Keywords

Autism Spectrum Disorder; Saliency; Eye Tracking; Semantics; Center Bias; Faces; Attention; Social Cognition

Corresponding author: Qi Zhao (eleqiz@nus.edu.sg). Department of Electrical and Computer Engineering, National University of Singapore, #E4-06-21, 4 Engineering Drive 3, 117583 Singapore. Phone: +65-6516-6658.

*Equal Contributions

Author Contributions

S.W., D.P.K., R.A. and Q.Z. designed experiments. S.W., M.J. and X.M.D. performed experiments. S.W., M.J. and Q.Z. analyzed data. E.A.L. helped with subject recruitment and assessment. S.W., R.A. and Q.Z. wrote the paper. All authors discussed the results and contributed toward the manuscript.

The authors declare no conflict of interest.

Introduction

People with autism spectrum disorder (ASD) show altered attention to, and preferences for, specific categories of visual information. When comparing social vs. non-social stimuli, individuals with autism show reduced attention to faces as well as to other social stimuli such as the human voice and hand gestures, but pay more attention to non-social objects (Dawson et al., 2005, Sasson et al., 2011), notably including gadgets, devices, vehicles, electronics, and other objects of idiosyncratic "special interest" (Kanner, 1943, South et al., 2005). Such atypical preferences are already evident early in infancy (Osterling and Dawson, 1994) and the circumscribed attentional patterns in eye tracking data can be found in 2–5 year-olds (Sasson et al., 2011) as well as in children and adolescents (Sasson et al., 2008). Several possibly related attentional differences are reported in children with ASD as well, including reduced social and joint attention behaviors (Osterling and Dawson, 1994) and orienting driven more by non-social contingencies rather than biological motion (Klin et al., 2009). We recently showed that people with ASD orient less towards socially relevant stimuli during visual search, a deficit that appeared independent of low-level visual properties of the stimuli (Wang et al., 2014). Taken together, these findings suggest that visual attention in people with ASD is driven by atypical saliency, especially in relation to stimuli that are usually considered socially salient, such as faces.

However, the vast majority of prior studies has used restricted or unnatural stimuli, e.g., faces and objects in isolation or even stimuli with only low-level features. There is a growing recognition that it is important to probe visual saliency with more natural stimuli (e.g., complex scenes taken with a natural background) (Itti et al., 1998, Parkhurst and Niebur, 2005, Cerf et al., 2009, Judd et al., 2009, Chikkerur et al., 2010, Freeth et al., 2011, Shen and Itti, 2012, Tseng et al., 2013, Xu et al., 2014), which have greater ecological validity and likely provide a better understanding of how attention is deployed in people with ASD when viewed in the real world (Ames and Fletcher-Watson, 2010). Although still relatively rare, natural scene viewing has been used to study attention in people with ASD, finding reduced attention to faces and the eye region of faces (Klin et al., 2002, Norbury et al., 2009, Riby and Hancock, 2009, Freeth et al., 2010, Riby et al., 2013), reduced attention to social scenes (Birmingham et al., 2011, Chawarska et al., 2013) and socially salient aspects of the scenes (Shic et al., 2011, Rice et al., 2012), and reduced attentional bias toward threat-related scenes when presented with pairs of emotional or neutral images (Santos et al., 2012). However, people with ASD seem to have similar attentional effects for animate objects as do controls when measured with a change detection task (New et al., 2010).

What is missing in all these prior studies is a comprehensive characterization of the various attributes of complex visual stimuli that could influence saliency. We aimed to address this issue in the present study, using natural scenes with rich semantic content to assess the spontaneous allocation of attention in a context closer to real-world free-viewing. Each scene included multiple dominant objects rather than a central dominant one, and we included both social and non-social objects, to allow direct investigation of the attributes that may differentially guide attention in ASD. Natural scene stimuli are less controlled, therefore requiring more sophisticated computational methods for analysis, along with a

larger sampling of different images. We therefore constructed a 3-layered saliency model with a principled vocabulary of pixel-, object-, and semantic-level attributes, quantified for all the features present in 700 different natural images (Xu et al., 2014). Furthermore, unlike previous work that focused on one or a few object categories with fixed prior hypotheses (Benson et al., 2009, Freeth et al., 2010, New et al., 2010, Santos et al., 2012), we used a data-driven approach free of assumptions that capitalized on using machine learning to provide an unbiased comparison among subject groups.

Results

People with ASD have higher saliency weights for low-level properties of images but lower weights for object- and semantic-based properties

Twenty people with ASD and nineteen controls who matched on age, IQ, gender, race and education (see **Experimental Procedures** and Table S1), freely viewed natural scene images for three seconds each (see **Experimental Procedures** for details). As can be seen qualitatively from the examples shown in Figure 1 (more examples in Figure S1), people with ASD made more fixations to the center of the images (Figure 1A–D), fixated on fewer objects when multiple similar objects were present in the image (Figure 1E, F), and seemed to have atypical preferences for particular objects in natural images (Figure 1G–L).

To formally quantify these phenomena and disentangle their contribution to the overall viewing pattern of people with ASD, we applied a computational saliency model with support vector machine (SVM) classifier to evaluate the contribution of five different factors in gaze allocation: (1) the image center, (2) the grouped pixel-level (color, intensity, and orientation), (3) object-level (size, complexity, convexity, solidity, and eccentricity), and (4) semantic-level (face, emotion, touched, gazed, motion, sound, smell, taste, touch, text, watchability, and operability; see Figure S2A for examples) features shown in each image, and (5) the background (i.e., regions without labeled objects) (see **Experimental Procedures** and Figure 2A for a schematic overview of the computational saliency model; see Table 1 for detailed description of features). Note that besides pixel-level features, each labeled object always had all object-level features and may have one or multiple semantic-level features (i.e., its semantic label(s)), while regions without labeled objects only had pixel-level features.

Our computational saliency model could predict fixation allocation with an area under the receiver operating characteristic (ROC) curve (AUC) score of 0.936±0.048 (mean±SD across 700 images) for people with ASD and 0.935±0.043 for controls (paired t-test, P=0.52; see Supplemental Results and Figure S2B, C), suggesting that all subsequent reported differences between the two groups could not be attributed to differences in model fit between the groups. Model fit was also in accordance with our prior work on an independent sample of subjects and a different model training procedure (Xu et al., 2014) (0.940±0.042; Supplemental Results, Figure S2B, C and Supplemental Discussion). The computational saliency model outputs a saliency weight for each feature, which represents the relative contribution of that feature to predict gaze allocation. As can be seen in Figure 2B, there was a large image center bias for both groups, a well-known effect (e.g., (Bindemann, 2010)). This was followed by effects driven by object- and semantic-level features. Note that before

training the SVM classifier, we z-scored the feature vector for each feature dimension by subtracting it from its mean and dividing it by its standard deviation. This assured that saliency weights could be compared, and were not confounded by possibly different dynamic ranges for different features.

Importantly, people with ASD had a significantly greater image center, background, and pixel-level bias, but a reduced object-level bias and semantic-level bias (see Figure 2B legend for statistical details). The ASD group did not have any greater variance in saliency weights compared to controls (one-tailed F-test; all Ps>0.94; significantly less variance for all features except pixel-level features; see Supplemental Discussion). Notably, when we controlled for individual differences in the duration of total valid eye tracking data (due to slight differences in blinks, etc.; Figure S2D–G), as well as for the Gaussian blob size for objects, and Gaussian map σ for analyzing the image center, we observed qualitatively the same results (Figure S3 and Supplemental Results), further assuring their robustness. Finally, we addressed the important issue that the different features in our model were necessarily intercorrelated to some extent. We used a leave-one-feature-out approach (Yoshida et al., 2012) that effectively isolates the non-redundant contribution of each feature by training the model each time with all but one feature from the full model ("minus-one" model). The obtained relative contribution of features with this approach was still consistent with the results shown in Figure 2B (Figure S3 and Supplemental Results), showing that our findings could not result from confounding correlations among features in our stimulus set. Note that the very first fixation in each trial was excluded from all analyses (see **Experimental Procedures**), since each trial began with a drift correction that required subjects to fixate on a dot at the very center of the image to begin with.

When fitting the model for each fixation individually, fixation-by-fixation analysis confirmed the above results and further revealed how the relative importance of each factor evolved over time (Figure 3). Over successive fixations, both subject groups weighted objects (Figure 3D) and semantics (Figure 3E) more, but low-level features (Figure 3A–C) less, suggesting that there was an increase in the role of top-down factors based on evaluating the meaning of the stimuli over time. This observation is consistent with previous findings that we initially use low-level features in the image to direct our eyes ("bottom-up attention"), but that scene understanding emerges as the dominant factor as viewing proceeds ("top-down attention") (Mannan et al., 2009, Xu et al., 2014). The decreasing influence of the image center over time resulted from exploration of the image with successive fixations (Zhao and Koch, 2011). Importantly, people with ASD showed less of an increase in the weight of object and semantic factors, compared to controls, resulting in increasing group differences over time (Figure 3D, E), and a similar but inverted group divergence for effects of image background, pixel-level saliency, and image centers (Figure 3A–C). Similar initial fixations were primarily driven by the large center bias for both groups, while the diverged later fixations were driven by object-based and semantic factors (note different y-axis scales in Figure 3).

Thus, these results show an atypically large saliency in favor of low-level properties of images (image center, background textures and pixel-level features) over object-based

properties (object and semantic features) in people with ASD. We further explore the differences in center bias and semantic attributes in the next sections.

People with ASD looked more at the image center even when there was no object

We examined whether the tendency to look at image center could be attributed to stimulus content. We first selected all images with no objects in the center 2° circular area, resulting in a total of 99 images. We then compared the total number of fixations in this area on these images. The ASD group had more than twice the number of fixations of the control group (ASD: 61.6±34.1; controls: 29.1±25.5; unpaired t-test, $t(37)=3.36$, $P=0.0018$, effect size in Hedges' g (standardized mean difference): $g=1.06$; permutation $P<0.001$). We further analyzed temporal differences and observed that in the early stage of the free-viewing (the first second), the difference was smaller (ASD: 14.5±10.2; controls: 7.74±8.42; $t(37)=2.23$, $P=0.032$, $g=0.70$; permutation $P=0.022$), while in the second second (ASD: 16.9±12.6; controls: 6.74±5.71; $t(37)=3.22$, $P=0.0027$, $g=1.010$; permutation $P<0.001$) and the third second (ASD: 30.3±16.6; controls: 14.6±13.9; $t(37)=3.18$, $P=0.0030$, $g=1.00$; permutation $P=0.008$), the difference was larger. In conclusion, these findings suggested that the stronger center bias in people with ASD could not be attributed to object distribution in the images.

The mean distance of all fixations from the image center was significantly smaller than in the control group (ASD: 5.63±0.56°; controls: 6.17±0.60°; unpaired t-test, $t(37)=-2.92$, $P=0.0059$, $g=-0.92$; permutation $P=0.010$), and there were more fixations at the image center than periphery compared to the control group (Fisher's exact test: $P<0.001$). As viewing proceeded, people with ASD also tended to return to fixating more central locations, as shown by decreasing distance to center for later fixations (Figure S4A). This re-centering was even evident in individual trials.

In exploratory analyses of individual differences, we found that subjects with higher Autism Quotient (AQ) scores (Pearson correlation; pooled ASD and controls: $r=-0.42$, $P=0.010$; ASD only: $r=-0.0064$, $p=0.98$; controls only: $r=-0.095$, $p=0.73$; the correlation was primarily driven by the group difference; Figure 4A) and lower Intelligence Quotient (IQ) (pooled ASD and controls: $r=0.33$, $P=0.039$; Figure 4B) looked closer to the image center, suggesting that stronger autistic traits and lower cognitive ability contribute to a more pronounced center bias. However, we found no correlation of center bias with age (pooled ASD and controls: $r=-0.15$, $P=0.38$; Figure 4C), arguing against any simple explanation due to motor slowing. We note that these correlations should be considered exploratory due to the small sample sizes at this stage.

Interestingly, we found that people with ASD had a smaller overall number of fixations, longer saccade durations, and reduced saccade velocity (Figure S4 and Supplemental Results), all consistent with a difficulty in shifting attention to other locations. This in turn might contribute to the stronger center bias in people with ASD.

Object fixations in ASD were also less well aligned with object centers (measured by average distance from the fixation to the centroid of the respective object region; ASD: 2.65±0.12° (mean ±SD); controls: 2.54±0.098°; unpaired t-test, $t(37)=3.28$, $P=0.0023$,

g=1.03; permutation P<0.001), indicating an atypical foveation of visual objects in people with ASD.

People with ASD looked less, and had longer latency towards, semantic features

We verified model-based results with a more standard fixation-based analysis (see Supplemental Results for full details). For effects at the semantic level, we compared fixation proportions, latency of the first fixation, and fixation duration (Figure 4D–F). People with ASD had fewer fixations on semantic features but more fixations on the background (Figure 4D), and they fixated semantic features significantly later than did the control group (Figure 4E). Although the total fixation duration per trial was significantly shorter in people with ASD, their mean duration per fixation was longer (Figure 4F). In conclusion, fixation analysis confirmed the results derived from our saliency model and showed stronger attention to the background rather than to semantic features in people with ASD.

Analysis of specific semantic categories

In Figure 3, we found that both groups showed a decreasing weighting of low-level features (center bias, background textures, and pixel-level attributes) but an increasing weighting of object and semantic attributes as viewing proceeded. One major advantage of our natural scene stimuli was that there was a broad range of different semantic categories that could be compared. Therefore, we next used the expanded semantic feature set (treating each of the twelve semantic attributes as a different channel) to train the saliency model (Figure 2A, Figure 5 and Figure S5). We found that motion, smell and touch features had significantly lower weights in people with ASD when training the model using all fixations (Figure 5). We also compared the evolution of saliency weights over serial order of fixations (Figure S5).

As shown in Figure S5, saliency weights of most semantic categories increased over time while low-level saliency weights decreased, confirming that semantic cues played a greater role as viewing proceeded. For most of the semantic-level features, the saliency weights for people with ASD were significantly lower than for controls, showing reduced attention to semantic features in ASD. Importantly, this difference only occurred at later fixations, consistent with the temporally increasing group differences in aggregate semantic weights we had shown earlier (Figure 3).

It is notable that the weights of face and emotion attributes were relatively high for initial fixations, suggesting that these attributes attracted attention more rapidly, an effect that could not be explained by a possible center bias for faces appearing in the images (see Figure S6A, B). We next examined in more detail the face and emotion attributes, two attributes that are at the focus of autism research.

We first observed that people with ASD had marginally reduced weights for faces (Figure 5; using all fixations: unpaired t-test, $t(37)=-1.71$, $P=0.095$, $g=-0.54$; permutation $P=0.088$; also see Figure S5E for fixation-by-fixation weights; see Figure 1G, H for examples) but not emotion ($t(37)=-0.042$, $P=0.97$, $g=-0.013$; permutation $P=0.99$), as well as a significant interaction (two-way repeated-measure ANOVA (subject group X semantic attribute of face

vs. emotion); main effect of subject group: $F(1,37)=1.91$, $P=0.17$, $\eta^2=0.017$; main effect of semantic attribute: $F(1,37)=120$, $P=3.42\times10^{-13}$, $\eta^2=0.48$; interaction: $F(1,37)=4.17$, $P=0.048$, $\eta^2=0.017$). The face attribute contained all faces (neutral and emotional) and the emotion attribute contained a subset of the faces from the face attribute with emotional expressions (all faces with the emotion attribute label also had the face attribute label by definition). These patterns thus indicate that people with ASD had reduced attention towards faces, regardless of facial emotions, an effect that became significant however only at later fixations (see Figure S5E).

Additionally, people with ASD had reduced attention to look at objects gazed upon by a human or animal in the scene (Figure S5G; Figure 5: $t(37)=-1.88$, $P=0.069$, $g=-0.60$; permutation $P=0.062$; see Figure 1I, J for examples), consistent with many other studies showing impaired joint attention in ASD (Mundy et al., 1994, Osterling and Dawson, 1994, Leekam and Ramsden, 2006, Brenner et al., 2007, Mundy et al., 2009, Freeth et al., 2010, Chevallier et al., 2012). However, compared to faces, people with ASD had disproportionately smaller attentional difference in written text, another highly salient cue in natural scenes (Cerf et al., 2009, Xu et al., 2014) (Supplemental Results).

Fixation-based analysis of semantic features

An examination of fixation density corroborated the above results. In Figure S6F, we can see that people with ASD had fewer fixations on all semantic features compared to controls (two-way repeated-measure ANOVA (subject group X semantic attribute); main effect of subject group: $F(1,407)=18.6$, $P=1.15\times10^{-4}$, $\eta^2=0.0035$; main effect of semantic attribute: $F(11,407)=732$, $P<10^{-20}$, $\eta^2=0.94$; interaction: $F(11,407)=1.64$, $P=0.084$, $\eta^2=0.0021$), especially for gazed, motion, taste, touch, text and watchability. Since different features had different sizes and the larger the feature, the more likely it was fixated, we analyzed maximum fixation density for each feature so as to minimize this size effect—as long as fixations were evenly spread out within features, the maximum density would be similar. People with ASD still showed smaller fixation densities for all semantic features (Figure S6G). Moreover, given that the strong fixation center bias could lead to more fixations on semantic features situated in the center of the image (Figure S6A, E), we discounted the center bias by applying an inverted Gaussian kernel (Figure S6D) to the fixation distribution, and found similar results (Figure S6H). Lastly, since the fixation distribution was intrinsically spatially biased (Figure S6E), we further removed this spatial bias by normalizing the fixation distribution to a spatially uniform distribution and then recomputed fixation proportion. Again, results were similar (Figure S6I). Statistical details of these control analyses are shown in Supplemental Results.

Since the computational saliency model was solely based on fixation density without incorporating any information of fixation latency and fixation duration, we next analyzed these two aspects through fixation-based analyses.

Both people with ASD and controls fixated on social attributes like faces (combined neutral and emotional) and emotion (emotional faces only) more rapidly than on other features (two-way repeated-measure ANOVA (subject group X semantic attribute); main effect of semantic attribute: $F(11,407)=93.3$, $P<10^{-20}$, $\eta^2=0.63$), consistent with their higher saliency

weights (more potent in attracting fixations). People with ASD had in general compatible latency with controls (main effect of subject group: $F(1,407)=0.039$, $P=0.85$, $\eta^2=9.48\times10^{-5}$), however, notably, compared to controls, people with ASD were significantly slower to fixate on face and emotion attributes, but faster to fixate on the non-social attributes of operability (natural or man-made tools used by holding or touching with hands) and touch (objects with a strong tactile feeling, e.g., a sharp knife, a fire, a soft pillow and a cold drink) (Figure S6J), consistent with some of the categories of circumscribed interests that have been reported in ASD (Lewis and Bodfish, 1998, Dawson et al., 2005, South et al., 2005, Sasson et al., 2011) (also see Figure 1K, L for higher fixation density on these attributes). The strong ANOVA interaction further confirmed the disproportionate latency difference between attributes ($F(11,407)=4.13$, $P=8.90\times10^{-6}$, $\eta^2=0.028$).

People with ASD had relatively longer mean duration per fixation for all semantic features (Figure S6K; two-way repeated-measure ANOVA (subject group X semantic attribute); main effect of subject group: $F(1,407)=2.67$, $P=0.11$, $\eta^2=0.042$), but both groups had the longest individual fixations on faces and emotion (main effect of semantic attribute: $F(11,407)=43.0$, $P<10^{-20}$, $\eta^2=0.20$; interaction: $F(11,407)=0.66$, $P=0.78$, $\eta^2=0.0031$). In particular, post-hoc t-tests revealed that people with ASD fixated on text significantly longer than did controls ($t(37)=2.85$, $P=0.0071$, $g=0.89$; permutation $P=0.006$).

These fixation-based additional analyses thus provide further detail to the roles of specific semantic categories. Whereas people with ASD were slower to fixate faces, they were faster to fixate mechanical objects, and had longer dwell times on text. These patterns are consistent with decreased attention to social stimuli, and increased attention to objects of special interest.

Interaction between pixel-, object- and semantic-level saliency

Due to the intrinsic spatial bias of fixations (e.g., center bias and object bias) and spatial correlations among features, we next conducted analyses to isolate the effect of each feature and examine the interplay between features in attracting fixations.

First, both subject groups had the highest saliency weight for faces (Figure 5) and the highest proportion of fixations on faces (Figure S6F). Could this semantic saliency weight pattern be explained by pixel-level or object-level features with which faces are correlated? We next computed pixel-level and object-level saliency for each semantic feature (see **Experimental Procedures**) and compared across semantic features. As can be seen from Figure S6C, neither pixel-level nor object-level saliency had the highest saliency for faces, nor the same pattern for all semantic features (Pearson correlation with semantic weight; pixel-level saliency: $r=0.088$, $P=0.79$ for ASD and $r=0.19$, $P=0.55$ for controls; object-level saliency: $r=0.20$, $P=0.53$ for ASD and $r=0.24$, $P=0.45$ for controls), indicating that semantic saliency was not in general simply reducible to pixel- or object-level saliency. Furthermore, center bias (occupation of center (Figure S6A): $r=0.43$, $P=0.16$ for ASD and $r=0.52$, $P=0.083$ for controls) and distribution of objects (distance to center (Figure S6B): $r=-0.067$, $P=0.84$ for ASD and $r=-0.0040$, $P=0.99$ for controls) could not explain semantic saliency either. In conclusion, our results argue that semantic saliency is largely independent of our set of low-level or object-level attributes.

Second, we examined the role of pixel-level saliency and object-level saliency when controlling for semantic saliency. For each semantic feature, we computed pixel-level saliency and object-level saliency for those semantic features that were most fixated (top 30% fixated objects across all images and all subjects) vs. least fixated (bottom 30% fixated objects across all images and all subjects). Since comparisons were made within the same semantic feature category, this analysis controlled semantic preference and could study the impact of pixel- and object-level saliency independently of semantic saliency.

We first explored two semantic features of interest—face and text. More fixated faces had both higher pixel-level saliency (Figure 6A; two-way repeated-measure ANOVA (subject group X object type); main effect of object type: $F(1,37)=109$, $P=1.38\times10^{-12}$, $\eta^2=0.66$) and object-level saliency (Figure 6B; main effect of object type: $F(1,37)=201$, $P=1.11\times10^{-16}$, $\eta^2=0.79$) than less fixated faces. Similarly, more fixated texts also had higher pixel-level saliency (Figure 6C; main effect of object type: $F(1,37)=609$, $P<10^{-20}$, $\eta^2=0.91$) and object-level saliency (Figure 6D; main effect of object type: $F(1,37)=374$, $P<10^{-20}$, $\eta^2=0.88$) than less fixated texts. These results suggested that both pixel-level saliency and object-level saliency contributed to attract more fixations to semantic features when controlling for semantic meanings. Interestingly, we found no difference between people with ASD and controls for all comparisons (main effect of subject group: all Ps>0.05; unpaired t-test: all Ps>0.05), suggesting that the different saliency weight (Figure 5) and fixation characteristics (Figure S6) of faces and text that we reported above between people with ASD and controls were not driven by pixel-level or object-level properties of faces and texts, but resulted from processes related to interpretation of the semantic meaning of those stimuli.

When we further analyzed the rest of the semantic features (Figure S7), we found that all features had reduced pixel-level saliency (main effect of object type: all Ps<0.01) and object-level saliency (all $Ps<10^{-4}$) for less fixated objects, confirming the role of pixel-level and object-level saliency in attracting attention. Again, we found no difference between people with ASD and controls for all comparisons (main effect of subject group: all Ps>0.05; unpaired t-test: all Ps>0.05) except gazed (Figure S7C; less fixated in ASD for pixel- and object-level saliency) and operability (Figure S7J; more fixated in ASD for object-level saliency only), suggesting that pixel-level and object-level saliency played a minimal role in reduced semantic saliency in people with ASD. This was further supported by no interaction between subject group and object type (all Ps>0.05 except for gazed and operability). Furthermore, we tried different definitions of more fixated and less fixated objects (e.g., top vs. bottom 10% fixated) and we found qualitatively the same results. Lastly, it is worth noting that the positive contribution of pixel-level saliency to semantic features does not conflict with its otherwise negative saliency weight (cf. Figure 3C) because (1) in the computational saliency model, all fixations were considered, including those on the background and other objects, and (2) the negative samples typically came from background textures instead of the less fixated semantic objects here (semantic objects mostly contained all positive samples) (see **Discussion** and Supplemental Discussion for further details).

In summary, we found that pixel-level and object-level saliency as well as center bias could not explain all of the saliency of semantic features, whereas even when controlling for

semantic saliency, pixel-level and object-level saliency was potent in attracting fixations. Importantly, neither pixel-level nor object-level saliency alone could explain the reduced semantic saliency that we found in ASD.

Discussion

In this study, we used natural scenes and a general data-driven computational saliency framework to study visual attention deployment in people with ASD. Our model showed that people with ASD had a stronger central fixation bias, stronger attention towards low-level saliency, and weaker attention towards semantic-level saliency. In particular, there was reduced attention to faces, and to objects of another's gaze, compared to controls, an effect that became statistically significant mainly at later fixations. The strong center bias in ASD was related to slower saccade velocity, but not fewer numbers of fixations nor object distribution. Furthermore, temporal analysis revealed that all attentional differences in people with ASD were most pronounced at later fixations, when semantic-level effects generally became more important. The results derived from the computational saliency model were further corroborated by direct analysis of fixation characteristics, which further revealed increased saliency for operability (i.e., mechanical and manipulable objects) and for text in ASD. We also found that the semantic saliency difference in ASD could not be explained solely by low-level or object-level saliency.

Possible caveats

Due to an overall spatial bias in fixations, and spatial correlations amongst object features, interactions between saliency weights are inevitable. For example, fixations tend to be on objects more often than on the background (Figure 4D), so pixel-level saliency will be coupled with object- and semantic-level saliency. If a fixated object has relatively lower pixel-level saliency than the background or unfixated objects, then the pixel-level saliency weights could be negative. Similarly, if the center region of an image has lower pixel-level saliency, center bias will lead to negative pixel-level saliency weights. To account for such interactions, we repeated our analysis by discounting the center bias using an inverted Gaussian kernel and by normalizing the spatial distribution of fixations. We also analyzed pixel-level and object-level saliency within a semantic feature category. It is worth noting that even when training the model with pixel-level features only (no object-level or semantic-level), the trained saliency weight of "intensity" was still negative for both groups, suggesting that subjects indeed fixated on some regions with lower pixel-level saliency and the negative weights were not computational artifacts of feature interactions.

It is important to keep in mind that, by and large, our images, as well as the selection and judgment of some of the semantic features annotated on them, were generated by people who do not have autism. That is, the photographs shown in the images themselves were presumably taken mostly by people who do not have autism (we do not know the details, of course). To some extent, it is thus possible that the stimuli and our analysis already builds in a bias, and would not be fully representative of how people with autism look at the world. There are two responses to this issue, and a clear direction for future studies. First, the large number of images drawn from an even larger set ensures wide heterogeneity; it is thus

highly likely that at least some images will correspond to familiar and preferred items for any given person, even though there are of course big individual differences across people in such familiarity and preference (this applies broadly to all people, not just to the comparison with autism). Second, there is in fact good reason to think that people with autism have generally similar experience, and also share many preferences, to typically developed individuals. That is, the case is not the same as if we were testing a secluded Amazonian tribe who has never seen many of the objects shown in our images. Our people with autism are all high-functioning individuals that live in our shared environment; they interact with the internet, all have cell phones, drive in cars, and so forth. Although there are differences (e.g., the ones we discover in this paper), they are sufficiently subtle that the general approach and set of images is still valid. Finally, these considerations suggest an obvious future experiment: have people with autism take digital photos of their environment to use as stimuli, and have people with autism annotate the semantic aspects of the images—a study beginning in our own laboratory.

We further discuss negative saliency weights, the difference between our current and previous model, behavioral variability in ASD, as well as center bias in the Supplemental Discussion.

Advantage of our stimuli, model and task

In this study, we used natural scene stimuli to probe saliency representation in people with ASD. Compared to most autism studies using more restricted stimulus sets, and/or more artificial stimuli, our natural scene stimuli offer a rich platform to study visual attention in autism under more ecologically relevant conditions (Ames and Fletcher-Watson, 2010). Furthermore, compared to previous studies which only focused on one or a few hypothesized categories like faces (Freeth et al., 2010) or certain scene types (Santos et al., 2012), our broad range of semantic objects in a variety of scene contexts (see Figure 1, Figure S1 and **Experimental Procedures**) offered a comprehensive sample of natural scene objects and we could thus readily compare the relative contribution of multiple features to visual attention abnormalities in people with ASD. Importantly, previous studies used either low-level stimuli or specific object categories but rarely studied their combined interactions or relative contributions to attention. One prior study showed that when examining fixations onto faces, pixel-level saliency does not differ between individuals with ASD and controls within the first five fixations (Freeth et al., 2011), consistent with our findings in the present study (see Figure 3C).

Furthermore, compared to studies with explicit top-down instructions (e.g. visual search tasks), the free-viewing paradigm used in the present study assesses the spontaneous allocation of attention in a context closer to real-world viewing conditions. We previously found that people with ASD have reduced attention to target-congruent objects in visual search, and that this abnormality is especially pronounced for faces (Wang et al., 2014). Other studies using natural scenes have found that people with ASD do not sample scenes according to top-down instructions (Benson et al., 2009), whereas one study reported normal attentional effects of animals and people in a scene in a change detection task (New et al., 2010). However, all these prior studies used a much smaller stimulus set than we did in the

present work, and none systematically investigated the effects of specific low-level and high-level factors as we do here.

Lastly, it is important to note that, while our results are of course relative to the stimulus set and the list of features we used, our selection of stimuli and features was unbiased with respect to hypotheses about ASD (identical to those in a prior study that was not about ASD at all; Xu et al., 2014). Similarly, the parameters used in our modeling were not in any way biased for hypotheses about ASD. Thus, the computational method that quantified the group differences we report could contribute to automated and data-driven classification and diagnosis for ASD, and aid in the identification of subtypes and outliers, as has been demonstrated already for some other disorders (Tseng et al., 2013).

Impaired attentional orienting in natural scenes

Previous work has reported deficits in orienting to both social and non-social stimuli in people with ASD (e.g., (Wang et al., 2014) and (Birmingham et al., 2011)) and increased autistic traits are associated with reduced social attention (Freeth et al., 2013). Studies have shown that while children with autism are able to allocate sustained attention (Garretson et al., 1990, Allen and Courchesne, 2001), they have difficulties in disengagement and shifting (Dawson et al., 1998, Swettenham et al., 1998). Our results likewise showed that, in natural scene viewing, people with ASD had longer dwell times on objects, a smaller overall number of fixations, longer saccade durations, and reduced saccade velocity, all consistent with a difficulty in shifting attention to other locations. Some of our stimuli contained multiple objects of the same category or with similar semantic properties (e.g., two cups in Figure 1E and two pictures in Figure 1F), but people with ASD tended to focus on only one of the objects rather than explore the entire image.

Altered saliency representation in ASD

In this study, we found reduced saliency for faces and gazed objects in ASD, consistent with prior work showing reduced attention to faces compared to inanimate objects (Dawson et al., 2005, Sasson et al., 2011). Given our spatial resolution, we did not analyze the features within faces, but it is known that the relative saliency of facial features is also altered in autism (Pelphrey et al., 2002, Neumann et al., 2006, Spezio et al., 2007, Kliemann et al., 2010). The atypical facial fixations are complemented by neuronal evidence of abnormal processing of information from the eye region of faces in blood-oxygen-level dependent (BOLD) fMRI (Kliemann et al., 2012) and in single cells recorded from the amygdala in neurosurgical patients with ASD (Rutishauser et al., 2013). It is thus possible that at least some of the reduced saliency for faces in ASD that we report in the present paper derived from an atypical saliency for the features within those faces.

We also report a reduced saliency towards gazed objects (objects in the image towards which people or animals in image are looking), consistent with the well-studied abnormal joint attention in ASD (Mundy et al., 1994, Osterling and Dawson, 1994, Leekam and Ramsden, 2006, Brenner et al., 2007, Mundy et al., 2009, Freeth et al., 2010, Chevallier et al., 2012) (see (Birmingham and Kingstone, 2009) for a review). Neuroimaging studies have shown that in autism, brain regions involved in gaze processing, such as the superior

temporal sulcus (STS) region, are not sensitive to intentions conveyed by observed gaze shifts (Pelphrey et al., 2005). By contrast, our fixation latency analysis revealed that people with ASD had faster saccades towards objects with the non-social feature of operability (mechanical or manipulable objects), consistent with increased valence rating on tools (esp. hammer, wrench, scissors and lock) (Sasson et al., 2012) and special interest in gadgets (South et al., 2005) in ASD. Thus, the decreased ASD saliency we found for faces and objects of shared attention, and the increased saliency for mechanical/manipulable objects, are quite consistent with what one would predict from the prior literature.

It remains an important further question to elucidate exactly what it is that is driving the saliency differences we report here. Saliency could arise from at least three separate factors: (1) low-level image properties (encapsulated in our pixel-wise saliency features), (2) reward value of objects (contributing to their semantic saliency weights), and (3) information value of objects (a less well understood factor that motivates people to look to locations where they expect to derive more information, such as aspects of the scene about which they are curious). An increased contribution of pixel-level saliency was apparent in our study, but was not the only factor contributing to altered attention in ASD. People with ASD have been reported to show a disproportionate impairment in learning based on social rewards (faces), compared to monetary rewards (Lin et al., 2012a) and have reduced preference for making donations to charities that benefit people (Lin et al., 2012b). This suggests that at least some of the semantic-level differences in saliency we report may derive from altered reward value for those semantic features in ASD. Future studies using instrumental learning tasks based on such semantic categories could further elucidate this issue (e.g., studies using faces, objects of shared attention, and mechanical objects as the outcomes in reward learning tasks).

Summary

In this comprehensive model-based study of visual saliency we found that (i) people with ASD look more at image centers, even when there is no object at the center. This may be due in part to slower overall saccade velocity. (ii) Temporal analyses showed that low-level saliency decreased but object- and semantic-level saliency increased over time for both groups. However, saliency weights diverged at later times, such that people with ASD fixated more on regions with pixel-level saliency, and less on regions with object-level and semantic-level saliency. (iii) People with ASD had atypical attention to specific semantic objects: they were slower to fixate on faces, but faster to fixate on mechanical and manipulable objects. (iv) Pixel- and object-level saliency could not explain the group differences in semantic saliency.

Experimental Procedures

Subjects

Twenty high-functioning people with ASD were recruited (Table S1). All ASD participants met DSM-IV/ICD-10 diagnostic criteria for autism, and all met the cutoff scores for ASD on the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 1989), and the Autism Diagnostic Interview-Revised (ADI-R) (LeCouteur et al., 1989) or Social Communication

Questionnaire (SCQ) when an informant was available (Table S1). We assessed IQ for participants using the Wechsler Abbreviated Scale of Intelligence (WASI™). The ASD group had a full scale IQ of 108.0±15.6 (mean±SD) and a mean age of 30.8±11.1 years.

Nineteen neurologically and psychiatrically healthy subjects with no family history of ASD were recruited as controls. Controls had a comparable full scale Intelligence Quotient (IQ) of 108.2±9.6 (t-test, P=0.95) and a comparable mean age of 32.3±10.4 years (t-test, P=0.66). Controls were also matched on gender, race and education. As expected, the ASD group had higher scores than controls in Social Responsiveness Scale-2 Adult Form Self Report (SRS-AR) (ASD: 83.8±18.5; control: 34.8±16.4; $P=8.46\times10^{-7}$) and Autism Spectrum Quotient (AQ) (ASD: 29.6±7.1; control: 15.2±4.8; $P=5.45\times10^{-8}$).

Subjects gave written informed consent and the experiments were approved by the Caltech Institutional Review Board. All subjects had normal or corrected-to-normal visual acuity. No enrolled subjects were excluded for any reasons.

Stimuli and task

We employed a free-viewing task with natural scene images from the OSIE dataset. This dataset has been characterized and described in detail previously (Xu et al., 2014). Briefly, the dataset contains 700 images, which have been quantified according to three pixel-level attributes (color, intensity, and orientation), five object-level attributes (size, complexity, convexity, solidity, and eccentricity), and twelve semantic attributes (face, emotion, touched, gazed, motion, sound, smell, taste, touch, text, watchability, and operability) annotated on a total of 5551 segmented objects. Since there are a large number and variety of objects in natural scenes, to make the ground truth data least dependent on subjective judgments, we followed several guidelines for the segmentation, as described in (Xu et al., 2014). Similar hand-labeled stimuli (Shen and Itti, 2012) have demonstrated advantages in understanding the saliency contributions from semantic features.

Images contain multiple dominant objects in a scene. The twelve semantic attributes fall into four categories: (i) directly relating to humans (i.e., face, emotion, touched, gazed); (ii) objects with implied motion in the image; (iii) relating to other (non-visual) senses of humans (i.e., sound, smell, taste, touch); and (iv) designed to attract attention or for interaction with humans (i.e., text, watchability, operability). The details of all attributes are described in Table 1 and some examples of semantic attributes are shown in Figure S2A.

Subjects viewed 700 images freely for three seconds each, in random orders. There was a drift correction before each trial. Images were randomly grouped into 7 blocks with each block containing 100 images. No trials were excluded.

Computational modeling and data analysis

We used support vector machine (SVM) classification to analyze the eye tracking data. We built a 3-layered architecture including pixel-, object-, and semantic-level features (see above). In addition, we included the image center and the background as features in our model to account for the strong image center effect in people with ASD. The SVM model was trained using the feature maps and the ground-truth human fixation maps, and generated

as output feature weights, which were linearly combined to best fit the human fixation maps. Thus, feature weights represented the relative contribution of each feature in predicting gaze allocation. A schematic flow chart of the model is detailed in Figure 2A. Importantly, separate models were trained individually (and hence saliency weights derived individually) for each individual subject, permitting statistical comparisons between ASD and control groups.

To compute the feature maps, we resized each image to 200 × 150 pixels. The pixel-level feature maps were generated using the well-known Itti-Koch saliency model (Itti et al., 1998), while the object- and semantic-level feature maps were generated by placing a 2D Gaussian blob (σ=2°) at each object's center. The Gaussian blobs only existed in the maps representing the corresponding attributes. The magnitude of the Gaussian was the calculated object-level or manually labeled semantic-level feature value.

To learn this model from the ground-truth human fixation maps (plotting all fixation points with a Gaussian blur, σ=1°), 100 pixels in each image were randomly sampled from the 10% most fixated regions as positive samples, and 300 pixels were sampled from the 30% least fixated regions as negative samples. All samples were normalized to have zero mean and unit variance in the feature space. Different from (Xu et al., 2014) where fixations were pooled from all subjects to generate a fixation map for model learning, in this work we learned one SVM model for each individual subject in order to statistically compare the attribute weights between people with ASD and controls.

In the saliency interaction analysis, pixel-level saliency for each object was selected as the maximum value of the object region in order to minimize the object size effect. This was because big objects tend to include uniform texture regions and thus have much smaller average pixel-level saliency, while fixations were normally attracted to the most salient region of an object. Thus, maximum saliency rather than average saliency was more representative of pixel-level saliency of an object. By definition, object-level saliency was computed as a single value for each object (Xu et al., 2014). Our center bias feature was defined as a Gaussian map (σ=1°) around the image center (Figure 2A).

In order to compare the model fit between people with ASD and controls, we also pooled all fixations for each group and used a subset of the data to train the model and a subset of data to test the model. Details of this model training and testing to compare model fit between groups are described in Supplemental Experimental Procedures.

In all analyses, we excluded the very first fixation since it was always in the center due to preceding drift correction. In fixation-by-fixation analyses, we included the subsequent first 10 fixations based on the average number of fixations for both groups. For trials with less than 10 fixations, we included data up to their last fixation, and thus there were fewer trials being averaged together for these later fixations.

Eye tracking, permutation, and fixation analyses methods are described in Supplemental Experimental Procedures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Elina Veytsman and Jessica Hopkins for help in recruiting research participants, Justin Lee and Tim Armstrong for collecting the data, Lynn Paul for psychological assessments, and Laurent Itti for valuable comments. This research was supported by a post-doctoral fellowship from the Autism Science Foundation (S.W.), a Fonds de Recherche du Québec en Nature et Technologies (FRQNT) predoctoral fellowship (X.M.D.), a National Institutes of Health Grant K99MH094409/R00MH094409 and NARSAD Young Investigator Award (D.P.K.), the Caltech Conte Center for the Neurobiology of Social Decision Making from NIMH and a grant from Simons Foundation (SFARI Award 346839, R.A.), and the Singapore Defense Innovative Research Program 9014100596 and the Singapore Ministry of Education Academic Research Fund Tier 2 MOE2014-T2-1-144 (Q.Z.).

References

Allen G, Courchesne E. Attention function and dysfunction in autism. Front Biosci. 2001; 6:D105–119. [PubMed: 11171544]

Ames C, Fletcher-Watson S. A review of methods in the study of attention in autism. Developmental Review. 2010; 30:52–73.

Benson V, Piper J, Fletcher-Watson S. Atypical saccadic scanning in autistic spectrum disorder. Neuropsychologia. 2009; 47:1178–1182. [PubMed: 19094999]

Bindemann M. Scene and screen center bias early eye movements in scene viewing. Vision Research. 2010; 50:2577–2587. [PubMed: 20732344]

Birmingham E, Cerf M, Adolphs R. Comparing social attention in autism and amygdala lesions: effects of stimulus and task condition. Social Neuroscience. 2011; 6:420–435. [PubMed: 21943103]

Birmingham E, Kingstone A. Human Social Attention. Annals of the New York Academy of Sciences. 2009; 1156:118–140. [PubMed: 19338506]

Brenner L, Turner K, Müller R-A. Eye Movement and Visual Search: Are There Elementary Abnormalities in Autism? J Autism Dev Disord. 2007; 37:1289–1309. [PubMed: 17120149]

Cerf M, Frady EP, Koch C. Faces and text attract gaze independent of the task: Experimental data and computer model. Journal of Vision. 2009; 9:10. [PubMed: 20053101]

Chawarska K, Macari S, Shic F. Decreased Spontaneous Attention to Social Scenes in 6-Month-Old Infants Later Diagnosed with Autism Spectrum Disorders. Biological Psychiatry. 2013; 74:195–203. [PubMed: 23313640]

Chevallier C, Kohls G, Troiani V, Brodkin ES, Schultz RT. The social motivation theory of autism. Trends in Cognitive Sciences. 2012; 16:231–239. [PubMed: 22425667]

Chikkerur S, Serre T, Tan C, Poggio T. What and where: A Bayesian inference theory of attention. Vision Research. 2010; 50:2233–2247. [PubMed: 20493206]

Dawson G, Meltzoff A, Osterling J, Rinaldi J, Brown E. Children with Autism Fail to Orient to Naturally Occurring Social Stimuli. J Autism Dev Disord. 1998; 28:479–485. [PubMed: 9932234]

Dawson G, Webb SJ, McPartland J. Understanding the Nature of Face Processing Impairment in Autism: Insights From Behavioral and Electrophysiological Studies. Developmental Neuropsychology. 2005; 27:403–424. [PubMed: 15843104]

Freeth M, Chapman P, Ropar D, Mitchell P. Do Gaze Cues in Complex Scenes Capture and Direct the Attention of High Functioning Adolescents with ASD? Evidence from Eye-tracking. J Autism Dev Disord. 2010; 40:534–547. [PubMed: 19904597]

Freeth M, Foulsham T, Chapman P. The influence of visual saliency on fixation patterns in individuals with Autism Spectrum Disorders. Neuropsychologia. 2011; 49:156–160. [PubMed: 21093466]

Freeth M, Foulsham T, Kingstone A. What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. PLoS ONE. 2013; 8:e53286. [PubMed: 23326407]

Garretson H, Fein D, Waterhouse L. Sustained attention in children with autism. J Autism Dev Disord. 1990; 20:101–114. [PubMed: 2324050]

Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell. 1998; 20:1254–1259.

Judd, T.; Ehinger, K.; Durand, F.; Torralba, A. Learning to predict where humans look. Computer Vision, 2009 IEEE 12th International Conference on; 2009. p. 2106-2113.

Kanner L. Autistic disturbances of affective contact. The Nervous Child. 1943; 2:217–250.

Kliemann D, Dziobek I, Hatri A, Baudewig J, Heekeren HR. The Role of the Amygdala in Atypical Gaze on Emotional Faces in Autism Spectrum Disorders. The Journal of Neuroscience. 2012; 32:9469–9476. [PubMed: 22787032]

Kliemann D, Dziobek I, Hatri A, Steimke R, Heekeren HR. Atypical Reflexive Gaze Patterns on Emotional Faces in Autism Spectrum Disorders. The Journal of Neuroscience. 2010; 30:12281–12287. [PubMed: 20844124]

Klin A, Jones W, Schultz R, Volkmar F, Cohen D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry. 2002; 59:809–816. [PubMed: 12215080]

Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. Nature. 2009; 459:257–261. [PubMed: 19329996]

LeCouteur A, Rutter M, Lord C. Autism diagnostic interview: A standardized investigator-based instrument. J Autism Dev Disord. 1989; 19:363–387. [PubMed: 2793783]

Leekam S, Ramsden CH. Dyadic Orienting and Joint Attention in Preschool Children with Autism. J Autism Dev Disord. 2006; 36:185–197. [PubMed: 16502142]

Lewis MH, Bodfish JW. Repetitive behavior disorders in autism. Mental Retardation and Developmental Disabilities Research Reviews. 1998; 4:80–89.

Lin A, Adolphs R, Rangel A. Impaired learning of social compared to monetary rewards in autism. Frontiers in Neuroscience. 2012a:6. [PubMed: 22347152]

Lin A, Tsai K, Rangel A, Adolphs R. Reduced social preferences in autism: evidence from charitable donations. Journal of Neurodevelopmental Disorders. 2012b; 4:8. [PubMed: 22958506]

Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L. Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. J Autism Dev Disord. 1989; 19:185–212. [PubMed: 2745388]

Mannan SK, Kennard C, Husain M. The role of visual salience in directing eye movements in visual object agnosia. Current Biology. 2009; 19:R247–R248. [PubMed: 19321139]

Mundy, P.; Sigman, M.; Kasari, C. The theory of mind and joint-attention deficits in autism. In: Baron-Cohen, S., et al., editors. Understanding other minds: Perspectives from autism. New York, NY, US: Oxford University Press; 1994. p. 181-203.

Mundy P, Sullivan L, Mastergeorge AM. A parallel and distributed-processing model of joint attention, social cognition and autism. Autism Research. 2009; 2:2–21. [PubMed: 19358304]

Neumann D, Spezio ML, Piven J, Adolphs R. Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. Social Cognitive and Affective Neuroscience. 2006; 1:194–202. [PubMed: 18985106]

New JJ, Schultz RT, Wolf J, Niehaus JL, Klin A, German TC, Scholl BJ. The scope of social attention deficits in autism: Prioritized orienting to people and animals in static natural scenes. Neuropsychologia. 2010; 48:51–59. [PubMed: 19686766]

Norbury CF, Brock J, Cragg L, Einav S, Griffiths H, Nation K. Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. Journal of Child Psychology and Psychiatry. 2009; 50:834–842. [PubMed: 19298477]

Osterling J, Dawson G. Early recognition of children with autism: A study of first birthday home videotapes. J Autism Dev Disord. 1994; 24:247–257. [PubMed: 8050980]

Parkhurst, D.; Niebur, E. Stimulus-driven guidance of visual attention in natural scenes. In: Itti, L., et al., editors. Neurobiology of Attention. Burlington, MA: Academic Press/Elsevier; 2005. p. 240-245.

Pelphrey K, Sasson N, Reznick JS, Paul G, Goldman B, Piven J. Visual Scanning of Faces in Autism. J Autism Dev Disord. 2002; 32:249–261. [PubMed: 12199131]

Pelphrey KA, Morris JP, McCarthy G. Neural basis of eye gaze processing deficits in autism. Brain. 2005; 128:1038–1048. [PubMed: 15758039]

Riby D, Hancock P, Jones N, Hanley M. Spontaneous and cued gaze-following in autism and Williams syndrome. Journal of Neurodevelopmental Disorders. 2013; 5:13. [PubMed: 23663405]

Riby D, Hancock PJB. Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. Journal of Intellectual Disability Research. 2009; 53:169–181. [PubMed: 19192099]

Rice K, Moriuchi JM, Jones W, Klin A. Parsing Heterogeneity in Autism Spectrum Disorders: Visual Scanning of Dynamic Social Scenes in School-Aged Children. Journal of the American Academy of Child & Adolescent Psychiatry. 2012; 51:238–248. [PubMed: 22365460]

Rutishauser U, Tudusciuc O, Wang S, Mamelak AN, Ross IB, Adolphs R. Single-Neuron Correlates of Atypical Face Processing in Autism. Neuron. 2013; 80:887–899. [PubMed: 24267649]

Santos A, Chaminade T, Da Fonseca D, Silva C, Rosset D, Deruelle C. Just Another Social Scene: Evidence for Decreased Attention to Negative Social Scenes in High-Functioning Autism. J Autism Dev Disord. 2012; 42:1790–1798. [PubMed: 22160371]

Sasson N, Dichter G, Bodfish J. Affective Responses by Adults with Autism Are Reduced to Social Images but Elevated to Images Related to Circumscribed Interests. PLoS ONE. 2012; 7:e42457. [PubMed: 22870328]

Sasson NJ, Elison JT, Turner-Brown LM, Dichter GS, Bodfish JW. Brief Report: Circumscribed Attention in Young Children with Autism. J Autism Dev Disord. 2011; 41:242–247. [PubMed: 20499147]

Sasson NJ, Turner-Brown LM, Holtzclaw TN, Lam KSL, Bodfish JW. Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. Autism Research. 2008; 1:31–42. [PubMed: 19360648]

Shen J, Itti L. Top-down influences on visual attention during listening are modulated by observer sex. Vision Research. 2012; 65:62–76. [PubMed: 22728922]

Shic F, Bradshaw J, Klin A, Scassellati B, Chawarska K. Limited activity monitoring in toddlers with autism spectrum disorder. Brain Research. 2011; 1380:246–254. [PubMed: 21129365]

South M, Ozonoff S, McMahon W. Repetitive Behavior Profiles in Asperger Syndrome and High-Functioning Autism. J Autism Dev Disord. 2005; 35:145–158. [PubMed: 15909401]

Spezio ML, Adolphs R, Hurley RSE, Piven J. Analysis of face gaze in autism using “Bubbles”. Neuropsychologia. 2007; 45:144–151. [PubMed: 16824559]

Swettenham J, Baron-Cohen S, Charman T, Cox A, Baird G, Drew A, Rees L, Wheelwright S. The Frequency and Distribution of Spontaneous Attention Shifts between Social and Nonsocial Stimuli in Autistic, Typically Developing, and Nonautistic Developmentally Delayed Infants. The Journal of Child Psychology and Psychiatry and Allied Disciplines. 1998; 39:747–753.

Tseng P-H, Cameron IM, Pari G, Reynolds J, Munoz D, Itti L. High-throughput classification of clinical populations from natural viewing eye movements. J Neurol. 2013; 260:275–284. [PubMed: 22926163]

Wang S, Xu J, Jiang M, Zhao Q, Hurlemann R, Adolphs R. Autism spectrum disorder, but not amygdala lesions, impairs social attention in visual search. Neuropsychologia. 2014; 63:259–274. [PubMed: 25218953]

Xu J, Jiang M, Wang S, Kankanhalli MS, Zhao Q. Predicting human gaze beyond pixels. Journal of Vision. 2014; 14:28. [PubMed: 24474825]

Yoshida M, Itti L, Berg David J, Ikeda T, Kato R, Takaura K, White Brian J, Munoz Douglas P, Isa T. Residual Attention Guidance in Blindsight Monkeys Watching Complex Natural Scenes. Current Biology. 2012; 22:1429–1434. [PubMed: 22748317]

Zhao Q, Koch C. Learning a saliency map using fixated locations in natural scenes. Journal of Vision. 2011; 11:9. [PubMed: 21393388]

Highlights

- A novel 3-layered saliency model with 5551 annotated natural scene semantic objects
- People with ASD have a stronger image center bias regardless of object distribution
- Generally increased pixel-level saliency but decreased semantic-level saliency in ASD
- Reduced saliency for faces and locations indicated by social gaze in ASD

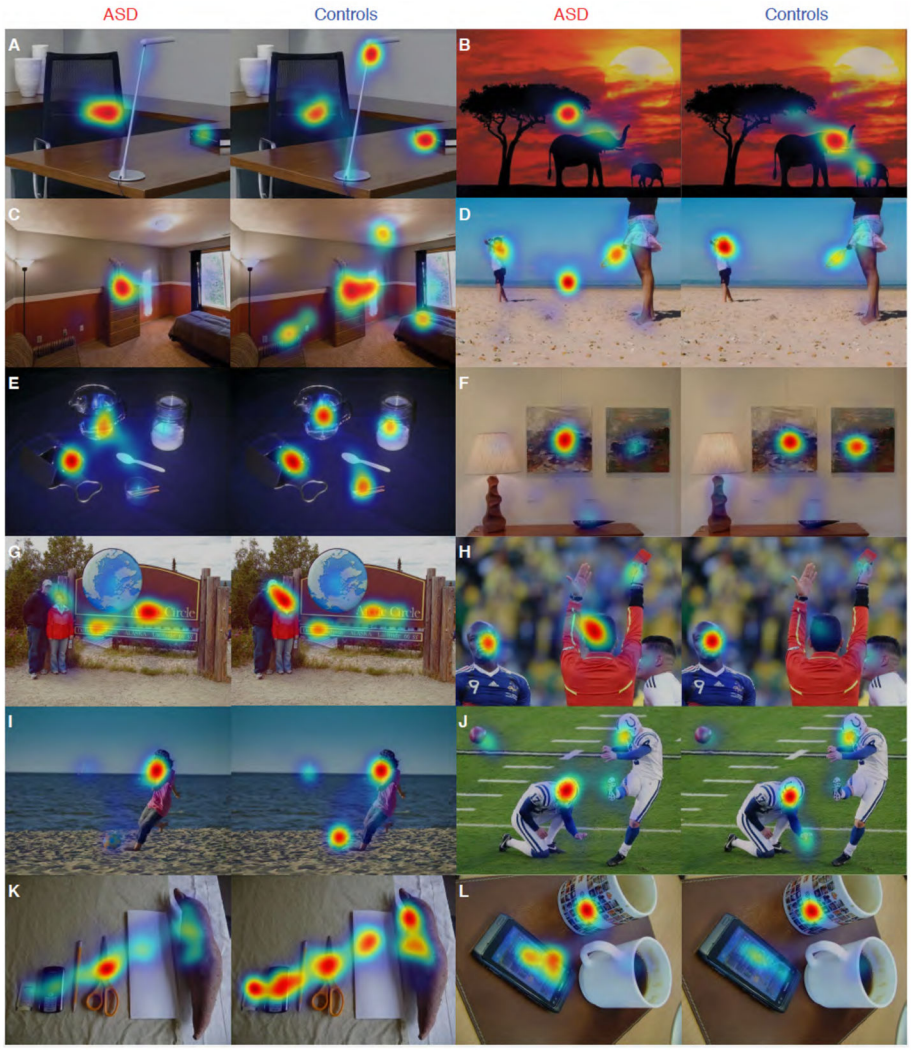

Figure 1.

Examples of natural scene stimuli and fixation densities from people with ASD (left) and controls (right). Heat map represents the fixation density. People with ASD allocated more fixations to the image centers **(A–D)**, fixated on fewer objects **(E, F)**, and had different semantic biases compared with controls **(G–L)**. See also Figure S1.

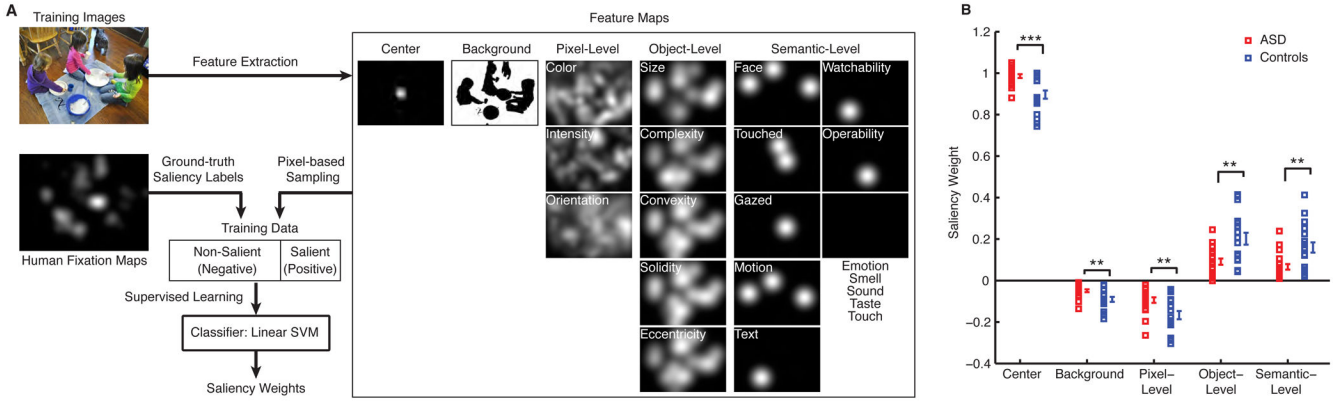

Figure 2.

Computational saliency model and saliency weights. **(A)** An overview of the computational saliency model. We applied a linear support vector machine (SVM) classifier to evaluate the contribution of five general factors in gaze allocation: the image center, the grouped pixel-level, object-level, and semantic-level features, and the background. Feature maps were extracted from the input images and included the three levels of features (pixel-, object-, and semantic-level) together with the image center and the background. We applied a pixel-based random sampling to collect the training data and trained on the ground-truth actual fixation data. The SVM classifier output were the saliency weights, which represented the relative importance of each feature in predicting gaze allocation. **(B)** Saliency weights of grouped features. People with ASD had a greater image center bias (ASD: 0.99±0.041 (mean±SD); controls: 0.90±0.086; unpaired t-test, $t(37)=4.18$, $P=1.72\times10^{-4}$, effect size in Hedges' g (standardized mean difference): $g=1.34$; permutation $P<0.001$), a relatively greater pixel-level bias (ASD: −0.094±0.060; controls: −0.17±0.087; $t(37)=3.06$, $P=0.0041$, $g=0.98$; permutation $P<0.001$) as well as background bias (ASD: −0.049±0.030; controls: −0.091±0.052; $t(37)=3.09$, $P=0.0038$, $g=0.99$; permutation $P=0.004$), but a reduced object-level bias (ASD: 0.091±0.067; controls: 0.20±0.13; $t(37)=-3.47$, $P=0.0014$, $g=-1.11$; permutation $P=0.002$) and semantic-level bias (ASD: 0.066±0.059; controls: 0.16±0.11; $t(37)=-3.37$, $P=0.0018$, $g=-1.08$; permutation $P=0.002$). Error bar denotes the standard error over the group of subjects. Asterisks indicate significant difference between people with ASD and controls using unpaired t-test. **: $P<0.01$, and ***: $P<0.001$. See also Figure S2, Figure S3 and Figure S4.

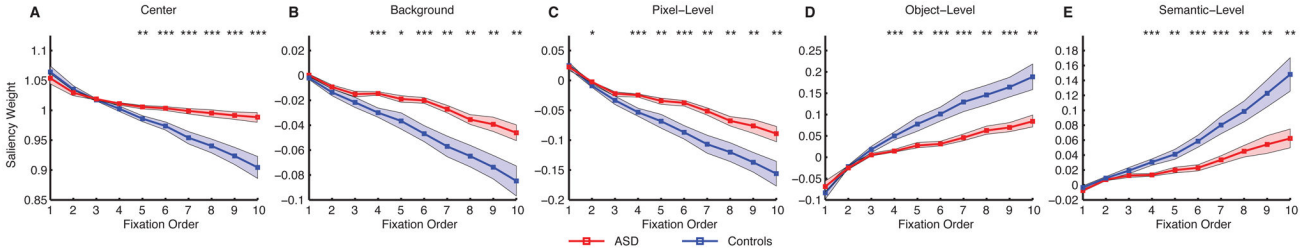

Figure 3.

Evolution of saliency weights of grouped features. Note that all data excluded the starting fixation, which was always at a fixation dot located at the image center; thus fixation number 1 shown in the figure is the first fixation away from the location of the fixation dot post stimulus onset. Shaded area denotes ±SEM over the group of subjects. Asterisks indicate significant difference between people with ASD and controls using unpaired t-test. *: P<0.05, **: P<0.01, and ***: P<0.001.

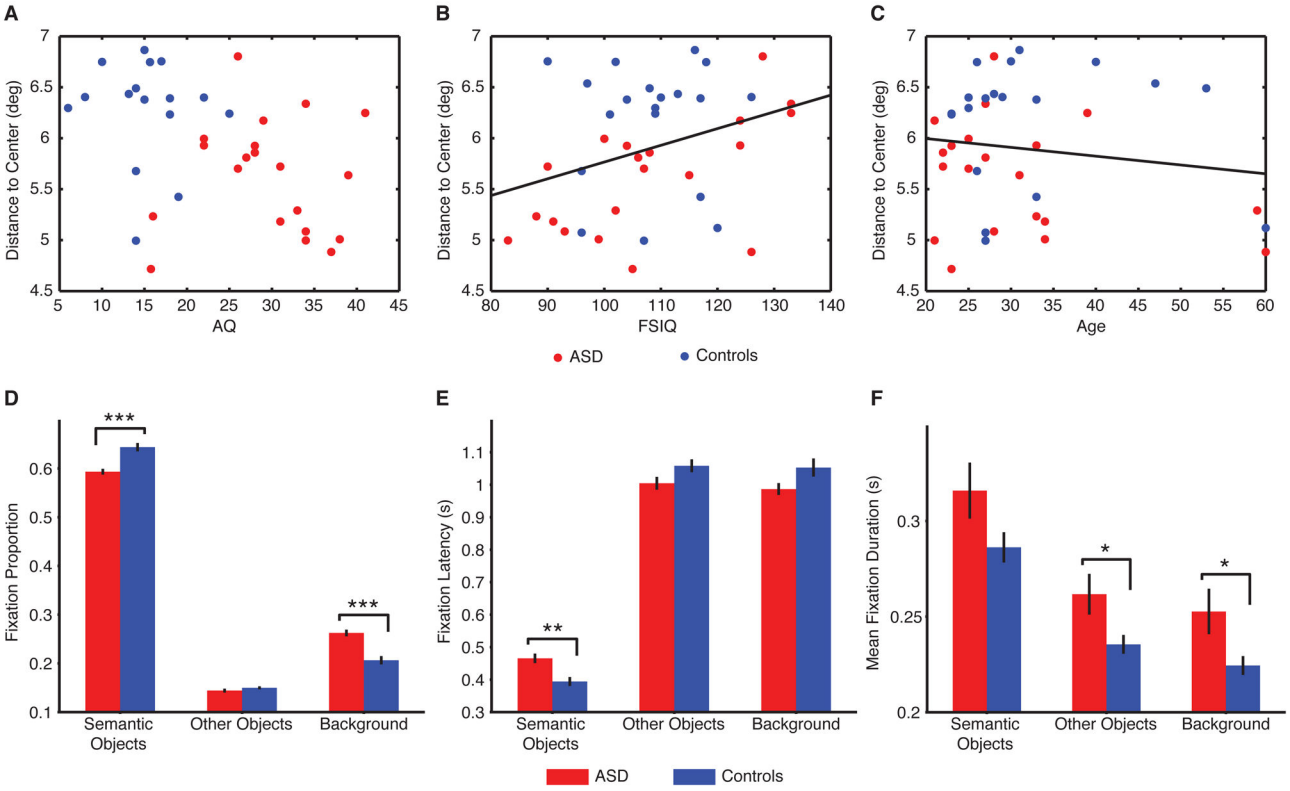

Figure 4.

Correlation and fixation analysis confirmed the results from our computational saliency model. **(A)** Correlation with AQ. **(B)** Correlation with FSIQ. **(C)** No correlation with age. Black lines represent best linear fit. Red: people with ASD. Blue: control. **(D)** People with ASD had fewer fixations on the semantic and other objects, but more on the background. **(E)** People with ASD fixated at the semantic objects significantly later than the control group, but not other objects. **(F)** People with ASD had longer individual fixations than controls, especially for fixations on background. Error bar denotes the standard error over the group of subjects. Asterisks indicate significant difference between people with ASD and controls using unpaired t-test. *: P<0.05, **: P<0.01, and ***: P<0.001.

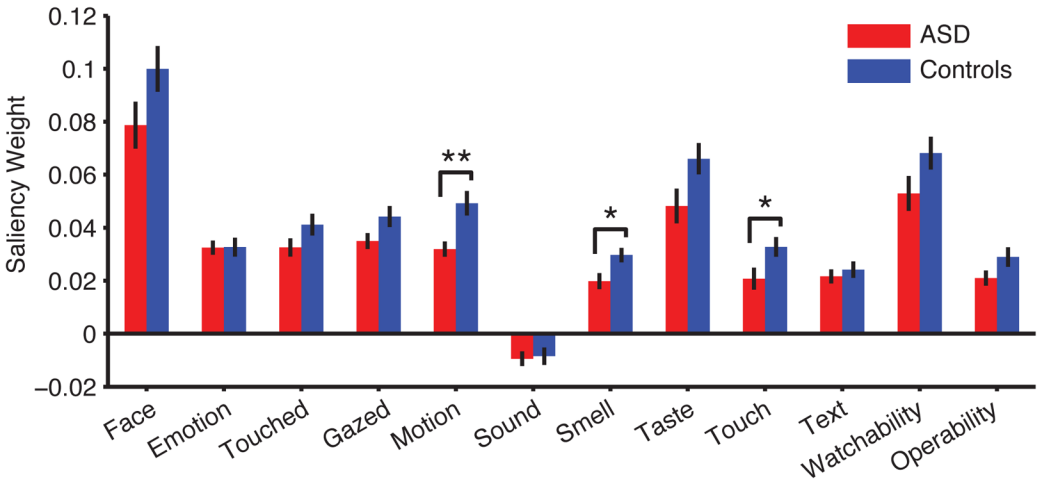

Figure 5.

Saliency weights of each of the twelve semantic features. We trained the classifier with the expanded full set of semantic features, rather than pooling over them (as in Figure 2B). Error bar denotes the standard error over the group of subjects. Asterisks indicate significant difference between people with ASD and controls using unpaired t-test. *: P<0.05, and **: P<0.01. See also Figure S5 and Figure S6.

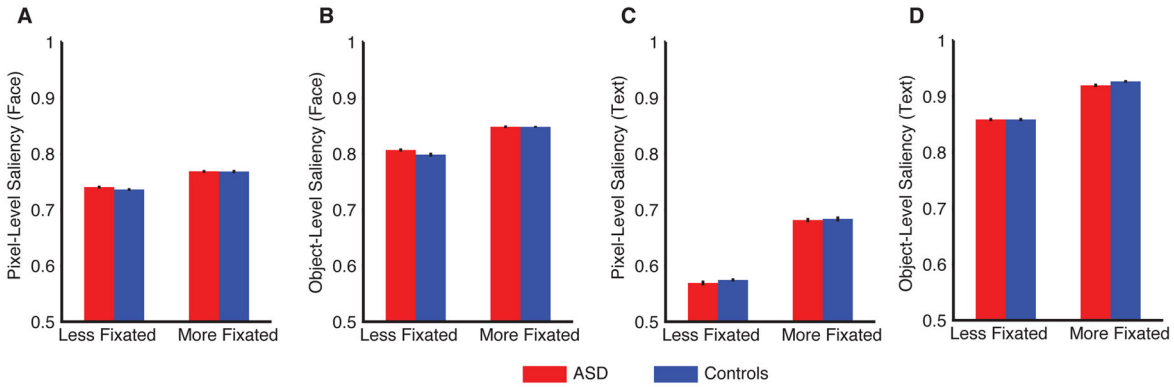

Figure 6.

Pixel- and object-level saliency for more vs. less fixated features. **(A)** Pixel-level saliency for more vs. less fixated faces. **(B)** Object-level saliency for more vs. less fixated faces. **(C)** Pixel-level saliency for more vs. less fixated texts. **(D)** Object-level saliency for more vs. less fixated texts. More fixated was defined as those 30% of faces/texts that were most fixated across all images and all subjects, and less fixated was defined as those 30% of faces/texts that were least fixated across all images and all subjects. Error bar denotes the standard error over objects. See also Figure S7.

Table 1

A summary of all features used in the computational saliency model.

Feature Type	Feature Name	Feature Description
	Center	A Gaussian map with σ=1°.
	Background	Regions without any labeled objects in the image.
Pixel-Level	**Color**	Color channel in the Itti-Koch model.
	Intensity	Intensity channel in the Itti-Koch model.
	Orientation	Orientation channel in the Itti-Koch model.
Object-Level	**Size**	The square root of the object's area.
	Complexity	The perimeter of the object's outer contour divided by the square root of its area.
	Convexity	The perimeter of the object's convex hull divided by the perimeter of its outer contour.
	Solidity	The area of the object divided by the area of its convex hull.
	Eccentricity	The eccentricity value of an ellipse that has the same second-moments as the object region.
Semantic-Level	**Face**	Back, profile, and frontal faces from human, animals and cartoons.
	Emotion	Faces from human, animals and cartoons with emotional expressions.
	Touched	Objects touched by a human or animal in the scene.
	Gazed	Objects gazed upon by a human or animal in the scene.
	Motion	Moving/flying objects, including humans/animals expressing meaningful gestures of postures that imply movement.
	Sound	Objects producing sound in the scene (e.g., a talking person, a musical instrument).
	Smell	Objects with a scent (e.g., a flower, a fish, a glass of wine).
	Taste	Food, drink, and anything that can be tasted.
	Touch	Objects with a strong tactile feeling (e.g., a sharp knife, a fire, a soft pillow, a cold drink).
	Text	Digits, letters, words, and sentences.
	Watchability	Man-made objects designed to be watched (e.g., a picture, a display screen, a traffic sign).
	Operability	Natural or man-made tools used by holding or touching with hands.
	Other	Objects labeled but not in any of the above categories