Individual differences in visual salience vary along semantic dimensions

Benjamin de Haas[a,b,1], Alexios L. Iakovidis[b], D. Samuel Schwarzkopf[b,c], and Karl R. Gegenfurtner[a]

[a]Department of Psychology, Justus Liebig Universität, 35394 Giessen, Germany; [b]Experimental Psychology, University College London, WC1H 0AP London, United Kingdom; and [c]School of Optometry & Vision Science, University of Auckland, 1142 Auckland, New Zealand

What determines where we look? Theories of attentional guidance hold that image features and task demands govern fixation behavior, while differences between observers are interpreted as a "noise-ceiling" that strictly limits predictability of fixations. However, recent twin studies suggest a genetic basis of gaze-trace similarity for a given stimulus. This leads to the question of how individuals differ in their gaze behavior and what may explain these differences. Here, we investigated the fixations of >100 human adults freely viewing a large set of complex scenes containing thousands of semantically annotated objects. We found systematic individual differences in fixation frequencies along six semantic stimulus dimensions. These differences were large (>twofold) and highly stable across images and time. Surprisingly, they also held for first fixations directed toward each image, commonly interpreted as "bottom-up" visual salience. Their perceptual relevance was documented by a correlation between individual face salience and face recognition skills. The set of reliable individual salience dimensions and their covariance pattern replicated across samples from three different countries, suggesting they reflect fundamental biological mechanisms of attention. Our findings show stable individual differences in salience along a set of fundamental semantic dimensions and that these differences have meaningful perceptual implications. Visual salience reflects features of the observer as well as the image.

visual salience | individual differences | eye movements

Humans constantly move their eyes (1). The foveated nature of the human visual system balances detailed representations with a large field of view. On the retina (2) and in the visual cortex (3) resources are heavily concentrated toward the central visual field, resulting in the inability to resolve peripheral clutter (4) and the need to fixate visual objects of interest. Where we move our eyes determines which objects and details we make out in a scene (5, 6).

Models of attentional guidance aim to predict which parts of an image will attract fixations based on image features (7–10) and task demands (11, 12). Classic salience models compute image discontinuities of low-level attributes, such as luminance, color, and orientation (13). These low-level models are inspired by "early" visual neurons and their output correlates with neural responses in subcortical (14) and cortical (15) areas thought to represent neural "salience maps." However, while these models work relatively well for impoverished stimuli, human gaze behavior toward richer scenes can be predicted at least as well by the locations of objects (16) and perceived meaning (9). When sematic object properties are taken into account, their weight for gaze prediction far exceeds that of low-level attributes (8, 17). A common thread of low- and high-level salience models is that they interpret salience as a property of the image and treat interindividual differences as unpredictable (7, 18), often using them as a "noise ceiling" for model evaluations (18).

However, even the earliest studies of fixation behavior noted considerable individual differences (19, 20), which recently gained wide-spread interest, ranging from behavioral genetics to computer science. Basic occulomotor traits, like mean saccadic amplitude and velocity, reliably vary between observers (21–28).

Gaze predictions based on artificial neural networks can improve when being trained on individual data (28, 29), or taking observer properties like age into account (30). The individual degree of visual exploration is correlated with trait curiosity (31, 32). Moreover, twin-studies show that social attention and gaze traces across complex scenes are highly heritable (33, 34). Taken together, these recent studies suggest that individual differences in fixation behavior are not random, but systematic. However, they largely focused on "content neutral" (32) or agnostic measures of gaze, like the spatial dispersion of fixations (32, 34, 35), the correlation of gaze traces (33), or the performance of individually trained models building on deep neural networks (28, 31, 29). Therefore, it remains largely unclear how individuals differ in their fixation behavior toward complex scenes and what may explain these differences. Here, we explicitly address this question: Can individual fixation behavior be explained by the systematic tendency to fixate different types of objects?

Specifically, we tested the hypothesis that individual gaze reflects individual salience differences along a limited number of semantic dimensions. We investigated the fixation behavior of >100 human adults (36) freely viewing 700 complex scenes, containing thousands of semantically annotated objects (8). We quantified salience differences as the individual proportion of cumulative fixation time or first fixations landing on objects with a given semantic attribute. In free viewing, the first fixations after image onset are thought to reflect "automatic" or "bottom-up"

Significance

To see clearly, humans have to constantly move their eyes and bring objects of interest to the center of gaze. Vision scientists have tried to understand this selection process, modelling the "typical observer." Here, we tested individual differences in the tendency to fixate different types of objects embedded in natural scenes. Fixation tendencies for faces, text, food, touched, and moving objects varied up to two- or threefold. These differences were present from the first eye movement after image onset and highly consistent across images and time. They predicted related perceptual skills and replicated across observer samples. This suggests individual gaze behavior is organized along semantic dimensions of biological significance, shaping the subjective way in which we see the world.

salience (37–39), especially for short saccadic latencies (40, 41). They may therefore reveal individual differences with a deep biological root. We tested the reliability of such differences across random subsets of images and across retests after several weeks. We also tested whether and to which degree individual salience models can improve the prediction of fixation behavior along these dimensions beyond the noise ceiling of generic models. To test the generalizability of salience differences, we replicated their set and covariance pattern across independent samples from three different countries. Finally, we explored whether individual salience differences are related to personality and perception, focusing on the example of face salience and face recognition skills for the latter.

Results

Reliable Salience Differences Along Semantic Dimensions. We tracked the gaze of healthy human adults freely viewing a broad range of images depicting complex everyday scenes (8). A first sample was tested at the University College London, United Kingdom (*Lon*; $n = 51$), and a replication sample at the University of Giessen, Germany (*Gi_1*; $n = 51$). This replication sample was also invited for a retest after 2 wk (*Gi_2*; $n = 48$). Additionally we reanalyzed a public dataset from Singapore [Xu et al. (8); $n = 15$].

First, we probed the individual tendency to fixate objects with a given semantic attribute, measuring duration-weighted fixations across a free-viewing period of 3 s. We considered a total of 12 semantic properties, which have previously been shown to carry more weight for predicting gaze behavior (on an aggregate group level) than geometric or pixel-level attributes (8). To test the consistency of individual salience differences across independent sets of images, we probed their reliability across 1,000 random (half-) splits of 700 images. Each random split was identical across all observers, and for each split individual differences seen

for one-half of the images were correlated with those seen for the other half. This way we tested the consistency of relative differences in fixation behavior across different subsets of images, without confounding them with image content (e.g., the absolute frequency of faces in a given subset of images). We found consistent individual salience differences ($r > 0.6$) for 6 of the 12 semantic attributes: Neutral *Faces*, *Emotion*al Faces, *Text*, objects being *Touched*, objects with a characteristic *Taste* (i.e., food and beverages), and objects with implied *Motion* (Fig. 1, gray scatter plots).

Observers showed up to twofold differences in the cumulative fixation time attracted by a given semantic attribute and the median consistency of individual differences across image splits for these six dimensions, ranged from $r = 0.64$ $P < 0.001$ (*Motion*) to $r = 0.94$, $P < 0.001$ (*Faces*; P values Bonferroni-corrected for 12 consistency correlations) (*SI Appendix*, Table S1, left hand side).

Previous studies have argued that extended viewing behavior is governed by cognitive factors, while first fixations toward a free-viewed image are governed by "bottom-up" salience (37–39), especially for short saccadic latencies (40, 41). Others have found that perceived meaning (9, 42) and semantic stimulus properties (8, 43) are important predictors of gaze behavior from the first fixation. We found consistent individual differences also in the proportion of first fixations directed toward each attribute. The range of individual differences in the proportion of first fixations directed to each of the six attributes was up to threefold, and thus even larger than that for cumulative fixation time. Importantly, these interobserver differences were consistent for all dimensions found for cumulative fixation time except *Motion* ($r = 0.34$, not significant), ranging from $r = 0.57$, $P < 0.001$ (*Taste*) to $r = 0.88$, $P < 0.001$ (*Faces*; P values Bonferroni-corrected for 12 consistency correlations) (green scatter plots in Fig. 1 and *SI Appendix*, Table S1, right hand side).
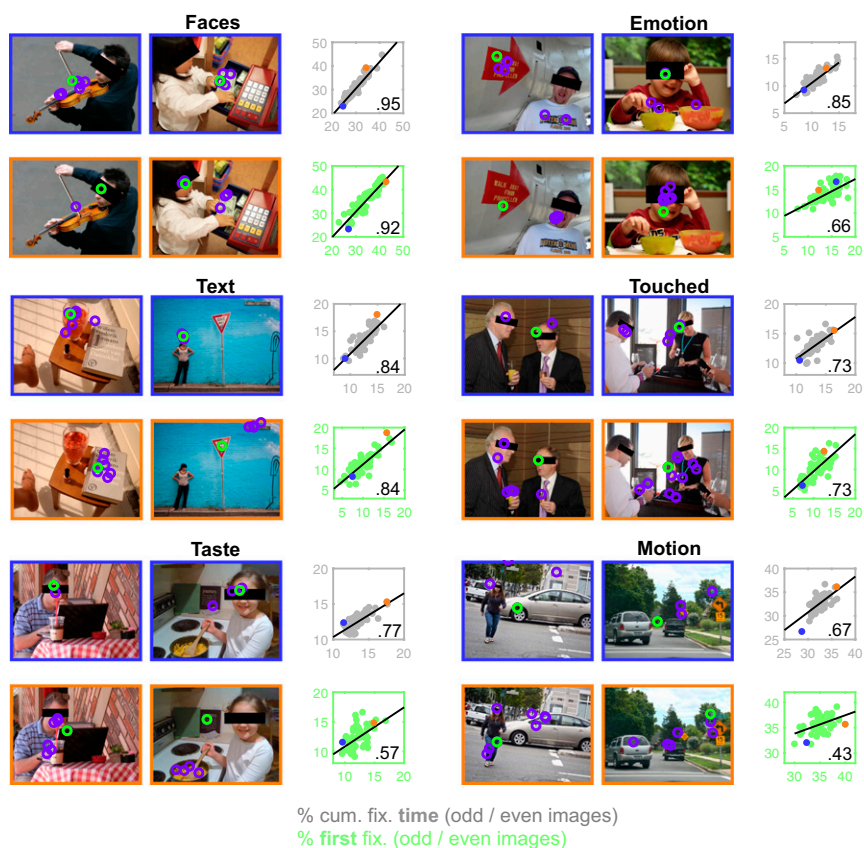

Fig. 1. Consistent individual differences in fixation behavior along six semantic dimensions. For each semantic attribute, the gray scatter plot shows individual proportions of cumulative fixation time for the odd versus even numbered images in the *Lon* dataset. The green scatter plot shows the corresponding individual proportions of first fixations after image onset. Black inset numbers give the corresponding Pearson correlation coefficient. For each dimension, two example images are given and overlaid with the fixations from one observer strongly attracted by the corresponding attribute (orange frames) and one observer weakly attracted by it (blue frames). The overlays show the first fixation after image onset as a green circle; any subsequent fixations are shown in purple. The two data points corresponding to the example observers are highlighted in the scatter plot, corresponding to the color of the respective image frames. All example stimuli from the OSIE dataset, published under the Massachusetts Institute of Technology license (8). Black bars were added to render faces unrecognizable for display purposes only (participants saw unmodified stimuli).

% cum. fix. **time** (odd / even images)
% **first** fix. (odd / even images)

de Haas et al.

These salience differences proved robust for different splits of images (Fig. 2A) and replicated across datasets from three different countries (*SI Appendix*, Fig. S1 and Table S1). For the confirmatory *Gi_1* dataset, we tested the same number of observers as in the original *Lon* set. A power analysis confirmed that this sample size yields >95% power to detect consistencies with a population effect size of $r > 0.5$. For cumulative fixation time (gray histograms in Fig. 2A), the six dimensions identified in the *Lon* sample, closely replicated in the *Gi_1*, *Gi_2* samples, as well as in a reanalysis of the public Xu et al. (8) dataset, with consistency correlations ranging from 0.65 (*Motion* in the *Gi_1* set) to 0.95 [*Faces* in the Xu et al. (8) dataset] (*SI Appendix*, Table S1, left column, and *SI Appendix*, Fig. S1). A similar pattern of consistency held for first fixations (green histograms in Fig. 2A), although the consistency correlation for *Emotion* missed statistical significance in the small Xu et al. dataset (8) (*SI Appendix*, Table S1, right column, and *SI Appendix*, Fig. S1).

The individual salience differences we found were consistent across subsets of diverse, complex images. To test whether they reflected stable observer traits, we additionally tested their retest reliability for the full image set across a period of 6–43 d (average 16 d; *Gi_1* and *Gi_2* datasets). Salience differences along all six semantic dimensions were highly consistent over time (Fig. 2B). This was true for both cumulative fixation time [retest reliabilities ranging from $r = 0.68$, $P < 0.001$ (*Motion*) to $r = 0.85$, $P < 0.001$ (*Faces*)] (gray bars in Fig. 2B and left column of *SI Appendix*, Table S1) and first fixations [retest reliabilities ranging from $r = 0.62$, $P < 0.001$ (*Taste*) to $r = 0.89$, $P < 0.001$ (*Text*)] (green bars in Fig. 2B and right column of *SI Appendix*, Table S1).

Additional control analyses confirmed that individual salience differences persisted independent of related visual field biases (*SI Appendix*, *Supplementary Results and Discussion* and Fig. S6).
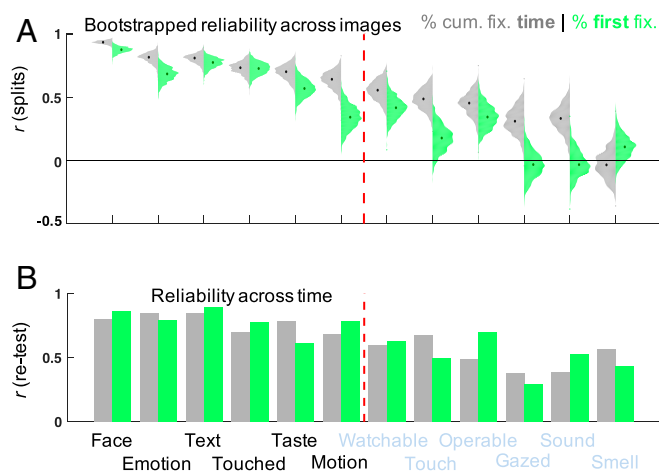
Fig. 2. Consistency of results across images and time. (*A*) Distribution of bootstrapped split-half correlations for each of the 12 semantic dimensions tested (as indicated by the labels on the *x* axis in *B*). The gray left-hand leaf of each distribution plot shows a histogram of split-half correlations for 1,000 random splits of the image set, the green right-hand leaf shows the corresponding histogram for first fixations after image onset. Overlaid dots indicate the median consistency correlation for each distribution. High split-half correlations indicate consistent individual differences in fixation across images for a given dimension. The dashed red line separates the six attributes found to be consistent dimensions of individual differences in the *Lon* sample. Data shown here is from the *Lon* sample and closely replicated across all datasets (*SI Appendix*, Fig. S1). (*B*) Retest reliability across the *Gi_1* and *Gi_2* samples. The magnitude of retest correlations for individual dwell time and proportion of first fixations is indicated by gray and green bars, respectively. All correlation and *P* values can be found in the *SI Appendix*, Table S1.

Individual Differences in Visual Exploration. Previous studies reported a relationship between trait curiosity and a tendency for visual exploration, as indexed by anticipatory saccades (31) or the dispersion of fixations across scene images (32). The latter was hypothesized to be a "content neutral" measure, independent of the type of salience differences we investigated here. Our data allows us to explicitly test this hypothesis. We ran an additional analysis, testing whether the number of objects fixated is truly independent of which objects an individual fixates preferentially.

First, we tested whether individual differences in visual exploration were reliable. The number of objects fixated significantly varied across observers, with a maximum/minimum ratio of 1.4 [Xu et al. (8)] to 1.9 (*Lon*) within a sample. Moreover, these individual differences were highly consistent across odd and even images in all four datasets (all $r > 0.98$, $P < 10^{-11}$) and showed good test-retest reliability ($r =, 80$, $P < 10^{-11}$ between *Gi_1* and *Gi_2*).

Crucially, however, we observed no significant relationship between the individual tendency for visual exploration and the proportion of first fixations landing on any of the six individual salience dimensions we identified (*SI Appendix*, Table S2, right hand side). For the proportions of cumulative dwell time, there was a moderate negative correlation between visual exploration and the tendency to fixate emotional expressions, which was statistically significant in the three bigger datasets (*Lon*, *Gi_1*, and *Gi_2*; all tests Holm-Bonferroni–corrected for six dimensions of interest) (*SI Appendix*, Table S2, left hand side). This negative correlation was not a mere artifact of longer dwelling on emotional expressions limiting the time to explore a greater number of objects. It still held when the individual proportion of dwell time on emotional expressions was correlated with the number of objects explored in images not containing emotional expressions ($r < −0.52$, $P < 0.001$ for all three datasets).

Individual Predictions Improve on the Generic Noise Ceiling. We took a first step toward evaluating how individual fixation predictions may improve on generic, group-based salience models. If individual differences were noise, then the mean of many observers should be the best possible predictor of individual gaze behavior. That is, the theoretical optimum of a generic model is the exact prediction of group fixation behavior for a set of test images, including fixation ratios along the six semantic dimensions identified above. Could individual predictions improve on this generic optimum?

We pooled fixation data across the 117 observers in the *Lon*, *Gi*, and Xu et al. (8) samples and randomly split the data into training and test sets of 350 images each (random splitting was repeated 1,000 times, with each set serving as test and training data once, totaling 2,000 folds). For each fold, we further separated a target individual from the remaining group, iterating through all individuals in a leave-one-observer-out fashion. For each fold and target observer, the empirical fixation ratios of the remaining group served as the (theoretical) ideal prediction of a generic salience model for the test images. We compared the prediction error for this ideal generic model to that of an individualized prediction.

The individual model was based on the assumption that fixation deviations from the group generalize from training to test data. It thus adjusted the prediction of the ideal generic model, based on the target individual's deviation from the group in the training data. Specifically, the target individual's fixation ratios for the training set were converted into units of SDs from the group mean. These z-scores were then used to predict individual fixation ratios for the test images, based on the mean and SD of the remaining group for the test set. Note that the individual model should perform worse than the ideal generic one if deviations from the group are random (see *SI Appendix*, *Supplementary Methods* for details).

Averaged across folds and cumulated across dimensions, the individual model reduced the prediction error for cumulative dwell time ratios for 89% of observers ($t_{116} = 11.39$, $P < 0.001$) and for first fixation ratios for 77% of observers ($t_{116} = 8.32$, $P < 0.001$).
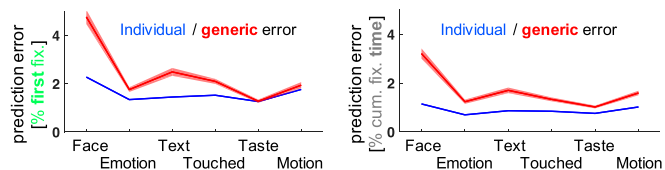
Fig. 3. Individual and generic prediction errors for fixation behavior. Prediction errors for proportions of fixations along the six semantic dimensions of individual salience. The (theoretical) ideal generic model predicted the group mean exactly, while individual models aimed to adjust predictions based on deviations from the group (seen for an independent set of training images). Prediction errors for the individual and generic models are shown in blue and red, as indicated. The line plots (shades) indicate the mean prediction error (±1 SEM) across observers. First fixation data shown on the *Left*, and cumulative dwell time on the *Right*, as indicated by the axis labels. For corresponding predictions and empirical data see *SI Appendix*, Fig. S2.

Across the group, this corresponded to a reduction of the mean cumulative prediction error from 10.09% (± 0.42% SEM) to 5.33% (± 0.10% SEM) for cumulative dwell time ratios and from 14.31% (± 0.56%, SEM) to 9.61% (± 0.10% SEM) for first fixation ratios. Individual predictions explained 74% of the error variance of ideal generic predictions for cumulative dwell time ratios and 58% of this error variance "beyond the noise ceiling" for first fixation ratios (again, averaged across folds and cumulated across dimensions) (see Fig. 3 and *SI Appendix*, Fig. S2 for individual dimensions).

Covariance Structure of Individual Differences in Semantic Salience. Having established reliable individual differences in fixation behavior along semantic dimensions, we further explored the space of these differences by quantifying the covariance between them. For this analysis we collapsed neutral and emotional faces into a single *Faces* label, because they are semantically related and corresponding differences were strongly correlated with each other ($r = 0.74$, $P < 0.001$; $r = 0.81$, $P < 0.001$ for cumulative fixation times and first fixations, respectively). Note that we decided to keep these two dimensions separated for the analyses above because the residuals of fixation times for emotional faces still varied consistently when controlling for neutral faces ($r = 0.73$, $P < 0.001$), indicating an independent component (however, the same was not true for first fixations, $r = 0.24$, not significant).

The resulting five dimensions showed a pattern of pairwise correlations that allowed the identification of two clusters (Fig. 4B). This was illustrated by the projection of the pairwise (dis)similarities onto a 2D space, using metric dimensional scaling (*SI Appendix*, Fig. S3). *Faces* and *Motion* were positively correlated with each other, but negatively with the remaining three attributes: *Text*, *Touched*, and *Taste*. Interestingly, *Faces*, the most prominent dimension of individual fixation behavior, was strongly anticorrelated with *Text* and *Touched*, the second and third most prominent dimensions [*Text*: $r = -0.62$, $P < 0.001$ and $r = -0.47$, $P < 0.001$ for cumulative fixation times and first fixations, respectively (Fig. 4 *A*, *Upper*); *Touched*: $r = -0.58$, $P < 0.001$ and $r = -0.80$, $P < 0.001$ (Fig. 4 *A*, *Lower*)]. These findings closely replicated across all four datasets (*SI Appendix*, Fig. S3). Pair-wise correlations between (z-converted) correlation matrices from different samples ranged from 0.68 to 0.95 for cumulative fixation times and from 0.91 to 0.98 for first fixations.

Perceptual Correlates of Salience Differences. If salience differences are indeed deeply rooted in the visual cortices of our observers, then this might have an effect on their perception of the world. We aimed to test this hypothesis by focusing on the most prominent dimension of salience differences: *Faces*, as indexed by the individual proportion of first fixations landing on faces (which is thought to be an indicator of bottom-up salience). Forty-six

observers from the *Gi* sample took the Cambridge Face Memory Test (CFMT) and we tested the correlation between individual face salience and face recognition skills. CFMT scores and the individual proportion of first fixations landing on faces correlated with $r = 0.41$, $P < 0.005$ (*SI Appendix*, Fig. S4, *Right*). Interestingly, this correlation did not hold for the individual proportion of total cumulative fixation time landing on faces, which likely represents more voluntary differences in viewing behavior ($r = 0.21$, not significant) (*SI Appendix*, Fig. S4, *Left*).

Additionally, we explored potential relationships with personality variables, but found no significant correlations between gaze behavior and standard questionnaire measures (*SI Appendix, Supplementary Results and Discussion* and Fig. S5).

Discussion

Individual differences in gaze traces have been documented since the earliest days of eye-tracking (19, 20). However, the nature of these differences was unclear, and therefore traditional salience models have either ignored them or used them as an upper limit for predictability ("noise-ceiling"). Our findings show that what was thought to be noise can actually be explained by a canonical set of semantic salience differences. These salience differences were highly consistent across hundreds of complex scenes, proved reliable in a retest after several weeks, and persisted independently of correlated visual field biases. This shows that visual salience is not just a factor of the image; individual salience differences are a stable trait of the observer, not only the set of these differences, but also their covariance structure replicated across independent samples from three different countries. This may partly be driven by environmental and image statistics (for example, faces are more likely to move than food). But it may also point to a neurobiological basis of these differences. This possibility is underscored by earlier studies showing that the visual salience of social stimuli is reduced in individuals with autism spectrum disorder (33, 44, 45). Most importantly, recent twin studies in infants and children show that individual differences in gaze traces are heritable (33, 34). The gaze trace dissimilarities investigated in these twin studies might be a manifestation of
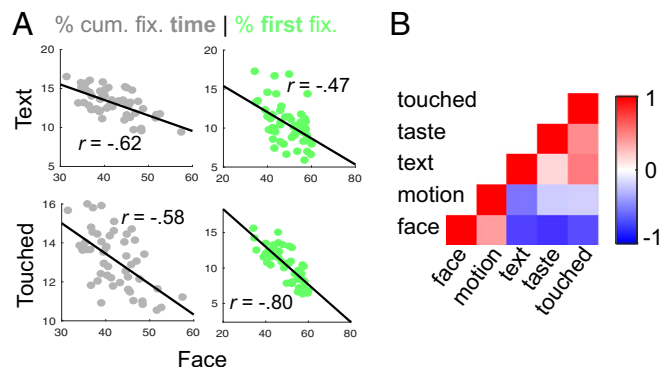
Fig. 4. Covariance of individual differences along semantic dimensions. (*A*) Gray scatter plots show the individual proportion of cumulative fixation time (in %) for *Faces* versus *Text* (*Left*) and *Faces* versus objects being *Touched* (*Right*). Green scatter plots show the corresponding data for the individual proportion of first fixations after image onset. (*B*) Correlation matrix for individual differences along five semantic dimensions (left hand side; note that the labels for emotional and neutral faces were collapsed for this analysis). Color indicates pairwise Pearson correlation coefficients as indicated by the bar. *Motion* and *Face* are positively correlated with each other, but negatively correlated with the remaining dimensions. This was also reflected by a two-cluster solution of metric dimensional scaling to two dimensions (*SI Appendix*, Fig. S3). All data shown are based on individual proportions of fixation time in the *Lon* dataset. For the corresponding consistency of this pattern for first fixations and across all four datasets, see *SI Appendix*, Fig. S3.

the salience differences we found here, which would imply a strong genetic component for individual salience differences.

Individual differences in gaze behavior have recently gained attention in fields ranging from computer science to behavioral genetics (28, 32–35). Previous findings converged to show such differences are systematic, but provided no clear picture of their nature. Our results show that individual salience varies along a set of semantic dimensions, which are among the best predictors of gaze behavior (8). Nevertheless, we cannot exclude the possibility of further dimensions of individual salience. For example, for some of the dimensions for which we found little or unreliable individual differences (watchable, touch, operable, gazed, sound, smell), such differences may have been harder to detect because they carry less salience overall (8). Lower overall numbers of fixations come with a higher risk of granularity problems. However, given that our dataset contains an average of over 5,000 fixations per observer for a wide range of images, it seems unlikely we missed any individual salience dimension of broad importance due to this problem. Future studies may probe the (unlabeled and potentially abstract) features of convolutional neural networks that carry weight for individual gaze predictions and may inform the search for further individual salience dimensions (17, 28).

Recent findings in macaque suggest that fixation tendencies toward faces and hands are linked to the development and prominence of corresponding domain-specific patches in the temporal cortex (46, 47). It is worth noting that most of the reliable dimensions of individual salience differences we found correspond to domain-specific patches of the ventral path [as is true for *Faces* (48–50), *Text* (51, 52), *Motion* (53), *Touched* (46, 54, 55), and maybe *Taste* (56)]. This opens the exciting possibility that these differences may be linked to neural tuning in the ventral stream.

Our findings raise important questions about the individual nature of visual perception. Two observers presented with the same image can end up with a different perception (5, 6) and interpretation (57) of this image when executing systematically different eye movements. Vision scientists may be chasing a phantom when "averaging out" individual differences to study the "typical observer" (58–60), and vice versa perception may be crucial to understanding individual differences in cognitive abilities (61, 62), personality (63, 64), social behavior (33, 44), clinical traits (65–67), and development (45).

We only took a first step toward investigating potential observer characteristics predicting individual salience here. Individual face salience was moderately correlated with face recognition skills. Interestingly, this was only true when considering the proportion of first fixations attracted by faces. Immediate saccades toward faces can have very short latencies and be under limited voluntary control (68, 69), likely reflecting bottom-up processing. This raises questions about the ontological interplay between face salience and recognition. Small initial differences may grow through mutual reinforcement of face fixations and superior perceptual processing, which would match the explanation of face processing difficulties in autism given by learning style theories (70).

We also investigated potential correlations with major personality dimensions, but found no evidence of such a relationship. Individual salience dimensions also appeared largely independent of the general tendency for visual exploration. However, one exception was the negative correlation between visual exploration and cumulative dwell time on emotional expressions, which may point to an anticorrelation of this salience dimension with trait curiosity (31, 32, 71). Future studies could use more comprehensive batteries to investigate the potential cognitive,

emotional, and personality correlates of individual salience. Individual salience may also be influenced by cultural differences (72), although it is worth noting that the space of individual differences we identified here seemed remarkably stable across culturally diverse samples. Finally, our experiments investigated individual salience differences for free viewing of complex scenes. Perceptual tasks can bias gaze behavior (11, 20) and diminish the importance of visual salience, especially in real world settings (12). It would be of great interest to investigate to which degree such differences persist in the face of tasks and whether they can affect task performance. For example, does individual salience predict attentional capture by a distractor, like the text of a billboard seen while driving?

In summary, we found a small set of semantic dimensions that span a space of individual differences in fixation behavior. These dimensions replicated across culturally diverse samples and also applied to the first fixations directed toward an image. Visual salience is not just a function of the image, but also of the individual observer.

Methods

Subjects, Materials, and Paradigm. The study comprised three original datasets [the *Lon* (n = 51), *Gi_1* (n = 51), and *Gi_2* (n = 48) samples (36)] and the reanalysis of a public dataset [the Xu et al. sample (8), n = 15]. The *Gi_2* sample was a retest of participants in the *Gi_1* dataset after an average of 16 d. The University College London Research Ethics Committee approved the *Lon* study and participants provided written informed consent. The Justus Liebig Universität Fb06 Local Ethics Committee (lokale Ethik-Kommission des Fachbereichs 06 der Justus Liebig Universität Giessen) approved the *Gi* study and participants provided written informed consent.

Participants in all samples freely viewed a collection of 700 complex everyday scenes, each shown on a computer screen for 3 s, while their gaze was tracked. Participants in the *Lon* sample additionally filled in standard personality questionnaires and participants in the *Gi_1* sample completed a standard test of face recognition skills (see *SI Appendix, Supplementary Methods* for more details).

Analyses. We harnessed preexisting metadata for objects embedded in the images (8) to quantify individual fixation tendencies for 12 semantic attributes. Specifically, we used two indices of individual salience: (*i*) the proportion of cumulative fixation time spent on a given attribute and (*ii*) the proportion of first fixations after image onset attracted by a given attribute (both expressed in percent) (see *SI Appendix, Supplementary Methods* for details).

We tested the split-half consistency of these measures across 1,000 random splits of images and their retest reliability across testing sessions of the *Gi_1* and *Gi_2* samples. Additionally, we tested whether individual deviations from the group mean for a set of training images could be used to predict individual group deviations for a set of test images and to which degree such individual predictions would improve on an ideal generic model. To investigate the covariance pattern of the six most reliable dimensions, we investigated the matrices of pairwise correlations and performed multidimensional scaling. The generalizability of the resulting pattern was tested as the correlation of similarity matrices across samples. We further tested pairwise correlations between dimensions of individual salience and personality in the *Lon* sample and between face salience and face recognitions skills in the *Gi* sample (see *SI Appendix, Supplementary Methods* for further details).

Availability of Data and Code. Anonymized fixation data and code to reproduce the results presented here are freely available at https://osf.io/n5v7t/.

1. K. R. Gegenfurtner, The interaction between vision and eye movements. *Perception* **45**, 1333–1357 (2016).
2. C. A. Curcio, K. A. Allen, Topography of ganglion cells in human retina. *J. Comp. Neurol.* **300**, 5–25 (1990).
3. R. F. Dougherty et al., Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J. Vis.* **3**, 586–598 (2003).
4. R. Rosenholtz, Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* **2**, 437–457 (2016).
5. J. M. Henderson, C. C. Williams, M. S. Castelhano, R. J. Falk, Eye movements and picture processing during recognition. *Percept Psychophys* **65**:725–734 (2003).
6. W. W. Nelson, G. R. Loftus, The functional visual field during picture viewing. *J. Exp. Psychol. Hum. Learn.* **6**, 391–399 (1980).

7. J. Harel, C. Koch, P. Perona, "Graph-based visual saliency" in *Proceedings of the 19th International Conference on Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2006), pp. 545–552.

8. J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, Q. Zhao, Predicting human gaze beyond pixels. *J. Vis.* **14**, 28 (2014).

9. J. M. Henderson, T. R. Hayes, Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav.* **1**, 743–747 (2017).

10. W. Einhäuser, M. Spain, P. Perona, Objects predict fixations better than early saliency. *J. Vis.* **8**, 18.1–18.26 (2008).

11. A. Borji, L. Itti, Defending Yarbus: Eye movements reveal observers' task. *J. Vis.* **14**, 29 (2014).

12. B. W. Tatler, M. M. Hayhoe, M. F. Land, D. H. Ballard, Eye guidance in natural vision: Reinterpreting salience. *J. Vis.* **11**, 5 (2011).

13. L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1254–1259 (1998).

14. B. J. White et al., Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nat. Commun.* **8**, 14263 (2017).

15. C. Bogler, S. Bode, J.-D. Haynes, Decoding successive computational stages of saliency processing. *Curr. Biol.* **21**, 1667–1671 (2011).

16. J. Stoll, M. Thrun, A. Nuthmann, W. Einhäuser, Overt attention in natural scenes: Objects dominate features. *Vision Res.* **107**, 36–48 (2015).

17. M. Kümmerer, T. S. A. Wallis, L. A. Gatys, M. Bethge, "Understanding low- and high-level contributions to fixation prediction" in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), pp. 4799–4808.

18. M. Kümmerer, T. S. A. Wallis, M. Bethge, Information-theoretic model comparison unifies saliency metrics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 16054–16059 (2015).

19. G. T. Buswell, *How People Look at Pictures: A Study of the Psychology and Perception in Art.* (Univ of Chicago Press, Oxford, England, 1935).

20. A. L. Yarbus, *Eye Movements During Perception of Complex Objects. Eye Movements and Vision* (Springer US, Boston, MA, 1967), pp. 171–211.

21. T. J. Andrews, D. M. Coppola, Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Res.* **39**, 2947–2953 (1999).

22. M. S. Castelhano, J. M. Henderson, Stable individual differences across images in human saccadic eye movements. *Can. J. Exp. Psychol.* **62**, 1–14 (2008).

23. J. M. Henderson, S. G. Luke, Stable individual differences in saccadic eye movements during reading, pseudoreading, scene viewing, and scene search. *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1390–1400 (2014).

24. I. Meyhöfer, K. Bertsch, M. Esser, U. Ettinger, Variance in saccadic eye movements reflects stable traits. *Psychophysiology* **53**, 566–578 (2016).

25. I. Rigas, O. V. Komogortsev, Current research in eye movement biometrics: An analysis based on BioEye 2015 competition. *Image Vis. Comput.* **58**, 129–141 (2017).

26. G. Bargary et al., Individual differences in human eye movements: An oculomotor signature? *Vision Res.* **141**, 157–169 (2017).

27. U. Ettinger et al., Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology* **40**, 620–628 (2003).

28. A. Li, Z. Chen, Personalized visual saliency: Individuality affects image perception. *IEEE Access* **6**, 16099–16109 (2018).

29. Y. Xu, S. Gao, J. Wu, N. Li, J. Yu, Personalized saliency and its prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* 10.1109/TPAMI.2018.2866563 (2018).

30. B. Yu, J. J. Clark, Personalization of saliency estimation. arXiv:1711.08000 (21 November 2017).

31. A. Baranes, P.-Y. Oudeyer, J. Gottlieb, Eye movements reveal epistemic curiosity in human observers. *Vision Res.* **117**, 81–90 (2015).

32. E. F. Risko, N. C. Anderson, S. Lanthier, A. Kingstone, Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition* **122**, 86–90 (2012).

33. J. N. Constantino et al., Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* **547**, 340–344 (2017).

34. D. P. Kennedy et al., Genetic influence on eye movements to complex scenes at short timescales. *Curr. Biol.* **27**, 3554–3560.e3 (2017).

35. M. Dorr, T. Martinetz, K. R. Gegenfurtner, E. Barth, Variability of eye movements when viewing dynamic natural scenes. *J. Vis.* **10**, 28 (2010).

36. B. de Haas, Individual differences in visual salience. Open Science Framework. https://osf.io/n5v7t/. Deposited 23 May 2018.

37. D. Parkhurst, K. Law, E. Niebur, Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* **42**, 107–123 (2002).

38. T. Foulsham, G. Underwood, What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *J. Vis.* **8**, 6.1–6.17 (2008).

39. W. Einhäuser, U. Rutishauser, C. Koch, Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J. Vis.* **8**, 2,1–2.19 (2008).

40. N. C. Anderson, E. Ort, W. Kruijne, M. Meeter, M. Donk, It depends on *when* you look at it: Salience influences eye movements in natural scene viewing and search early in time. *J. Vis.* **15**, 9 (2015).

41. M. Mackay, M. Cerf, C. Koch, Evidence for two distinct mechanisms directing gaze in natural scenes. *J. Vis.* **12**, 9 (2012).

42. J. M. Henderson, T. R. Hayes, Meaning guides attention in real-world scene images: Evidence from eye movements and meaning maps. *J. Vis.* **18**, 10 (2018).

43. M. Nyström, K. Holmqvist, Semantic override of low-level features in image viewing—Both initially and overall. *J. Eye Mov. Res.* **2**, 2:1–2:11 (2008).

44. S. Wang et al., Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. *Neuron* **88**, 604–616 (2015).

45. W. Jones, A. Klin, Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* **504**, 427–431 (2013).

46. M. J. Arcaro, P. F. Schade, J. L. Vincent, C. R. Ponce, M. S. Livingstone, Seeing faces is necessary for face-domain formation. *Nat. Neurosci.* **20**, 1404–1412 (2017).

47. K. Vinken, R. Vogels, A behavioral face preference deficit in a monkey with an incomplete face patch system. *Neuroimage* **189**, 415–424 (2019).

48. N. Kanwisher, G. Yovel, The fusiform face area: A cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 2109–2128 (2006).

49. D. Y. Tsao, S. Moeller, W. A. Freiwald, Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19514–19519 (2008).

50. K. Grill-Spector, K. S. Weiner, The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* **15**, 536–548 (2014).

51. B. D. McCandliss, L. Cohen, S. Dehaene, The visual word form area: Expertise for reading in the fusiform gyrus. *Trends Cogn. Sci. (Regul. Ed.)* **7**, 293–299 (2003).

52. S. Dehaene et al., How learning to read changes the cortical networks for vision and language. *Science* **330**, 1359–1364 (2010).

53. Z. Kourtzi, N. Kanwisher, Activation in human MT/MST by static images with implied motion. *J. Cogn. Neurosci.* **12**, 48–55 (2000).

54. T. Orlov, T. R. Makin, E. Zohary, Topographic representation of the human body in the occipitotemporal cortex. *Neuron* **68**, 586–600 (2010).

55. K. S. Weiner, K. Grill-Spector, Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychol. Res.* **77**, 74–97 (2013).

56. K. Adamson, V. Troiani, Distinct and overlapping fusiform activation to faces and food. *Neuroimage* **174**, 393–406 (2018).

57. J. C. Bush, P. C. Pantelis, X. Morin Duchesne, S. A. Kagemann, D. P. Kennedy, Viewing complex, dynamic scenes "through the eyes" of another person: The gaze-replay paradigm. *PLoS One* **10**, e0134347 (2015).

58. J. B. Wilmer, How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spat. Vis.* **21**, 561–579 (2008).

59. I. Charest, N. Kriegeskorte, The brain of the beholder: Honouring individual representational idiosyncrasies. *Lang. Cogn. Neurosci.* **30**, 367–379 (2015).

60. D. Peterzell, Discovering sensory processes using individual differences: A review and factor analytic manifesto. *Electron Imaging*, 1–11 (2016).

61. J. Haldemann, C. Stauffer, S. Troche, T. Rammsayer, Processing visual temporal information and its relationship to psychometric intelligence. *J. Individ. Differ.* **32**, 181–188 (2011).

62. T. R. Hayes, J. M. Henderson, Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *J. Vis.* **17**, 23 (2017).

63. S. J. Troche, T. H. Rammsayer, Attentional blink and impulsiveness: Evidence for higher functional impulsivity in non-blinkers compared to blinkers. *Cogn. Process.* **14**, 273–281 (2013).

64. D. W.-L. Wu, W. F. Bischof, N. C. Anderson, T. Jakobsen, A. Kingstone, The influence of personality on social attention. *Pers. Individ. Dif.* **60**, 25–29 (2014).

65. T. R. Hayes, J. M. Henderson, Scan patterns during scene viewing predict individual differences in clinical traits in a normative sample. *PLoS One* **13**, e0196654 (2018).

66. T. Armstrong, B. O. Olatunji, Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin. Psychol. Rev.* **32**, 704–723 (2012).

67. R. J. Molitor, P. C. Ko, B. A. Ally, Eye movements in Alzheimer's disease. *J. Alzheimer's Dis.* **44**, 1–12 (2015).

68. S. M. Crouzet, H. Kirchner, S. J. Thorpe, Fast saccades toward faces: Face detection in just 100 ms. *J. Vis.* **10**, 16.1–16.17 (2010).

69. L. Rösler, A. End, M. Gamer, Orienting towards social features in naturalistic scenes is reflexive. *PLoS One* **12**, e0182037 (2017).

70. N. Qian, R. M. Lipkin, A learning-style theory for understanding autistic behaviors. *Front. Hum. Neurosci.* **5**, 77 (2011).

71. S. Hoppe, T. Loetscher, S. Morey, A. Bulling, "Recognition of curiosity using eye movement analysis" in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers–UbiComp '15* (ACM Press, New York, 2015), pp. 185–188.

72. H. F. Chua, J. E. Boland, R. E. Nisbett, Cultural variation in eye movements during scene perception. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12629–12633 (2005).