Eye movements reveal epistemic curiosity in human observers

Adrien Baranes, Pierre-Yves Oudeyer, Jacqueline Gottlieb

Vision Research 117 (2015) 81–90

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Eye movements reveal epistemic curiosity in human observers

Adrien Baranes [a,*], Pierre-Yves Oudeyer [c,d], Jacqueline Gottlieb [a,b]

[a] Department of Neuroscience, Columbia University, United States
[b] The Kavli Institute for Brain Science, Columbia University, United States
[c] Inria, France
[d] Ensta ParisTech, France

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form 18 October 2015
Accepted 19 October 2015
Available online 12 November 2015

Keywords:
Saccades
Curiosity
Anticipation
Data mining
Random forests
Trivia questions

ABSTRACT

Saccadic (rapid) eye movements are primary means by which humans and non-human primates sample visual information. However, while saccadic decisions are intensively investigated in instrumental contexts where saccades guide subsequent actions, it is largely unknown how they may be influenced by curiosity – the intrinsic desire to learn. While saccades are sensitive to visual novelty and visual surprise, no study has examined their relation to epistemic curiosity – interest in symbolic, semantic information. To investigate this question, we tracked the eye movements of human observers while they read trivia questions and, after a brief delay, were visually given the answer. We show that higher curiosity was associated with earlier anticipatory orienting of gaze toward the answer location without changes in other metrics of saccades or fixations, and that these influences were distinct from those produced by variations in confidence and surprise. Across subjects, the enhancement of anticipatory gaze was correlated with measures of trait curiosity from personality questionnaires. Finally, a machine learning algorithm could predict curiosity in a cross-subject manner, relying primarily on statistical features of the gaze position before the answer onset and independently of covariations in confidence or surprise, suggesting potential practical applications for educational technologies, recommender systems and research in cognitive sciences. With this article, we provide full access to the annotated database allowing readers to reproduce the results. Epistemic curiosity produces specific effects on oculomotor anticipation that can be used to read out curiosity states.

1. Introduction

Curiosity is defined as the intrinsic motivation to learn and acquire information, and plays a central role in intelligent behavior including in development, learning and exploration (Berlyne, 1954; Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Oudeyer, Baranes, & Kaplan, 2013). Psychological theories formulated in the 1960s and 1970s distinguished between perceptual curiosity – a desire to obtain new sensory inputs – and epistemic curiosity – an interest in new knowledge or semantic information (Lowenstein, 1994). More recently, epistemic curiosity was associated with cortical and subcortical structures in human observers, including activation of reward-related structures (Kang et al., 2009), and memory enhancement through reward modulations of hippocampal mechanisms (Gruber, Gelman, & Ranganath, 2014).

An open question however concerns the links between curiosity and selective attention. Attention, along with working memory, is critical for learning and selective information processing (Cardoso-Leite & Bavelier, 2014; Gottlieb et al., 2013). In humans and non-human primates, visual attention and rapid eye movements (saccades) are the primary means by which subjects sample visual information, and are sensitive to value and motivation (Gottlieb, 2012; Gottlieb, Hayhoe, Hikosaka, & Rangel, 2014; Tatler, Hayhoe, Land, & Ballard, 2011). While a recent study has shown that personality measures of trait curiosity correlate with the numbers of saccades and numbers of regions explored during free-viewing of complex scenes (Risko, Anderson, Lanthier, & Kingstone, 2012), nothing is known about the links between eye movements and epistemic curiosity – interest in semantic information.

In this report we examined this question by tracking the eye movements of human observers while they were presented a series of trivia questions that created high or low epistemic curiosity states. Because curiosity can covary with other epistemic factors such as confidence and surprise, we asked subjects to provide

* Corresponding author at: Department of Neuroscience, Columbia University, 1051 Riverside Drive, Kolb Research Annex, New York, NY 10032, United States.
E-mail address: Adrien.baranes@gmail.com (A. Baranes).

http://dx.doi.org/10.1016/j.visres.2015.10.009
0042-6989/© 2015 The Authors. Published by Elsevier Ltd.

independent ratings of the 3 subjective states. We tested the hypothesis that curiosity will influence eye movement control, and that these influences would be sufficiently specific to allow curiosity to be "read out" from eye movements using data mining algorithms. The results confirmed both predictions. We show that curiosity enhanced anticipatory eye movements toward the expected location of the answer and the dwell time on the answer after it was presented, without affecting other metrics of saccades or fixations. The ocular signatures of high or low curiosity, confidence or surprise were sufficiently specific so that a machine learning algorithm could discriminate these levels with above-chance accuracy across multiple individual observers.

2. Methods

2.1. Subjects

Twenty subjects (11 women) were recruited from the Columbia University community and were compensated for their participation at the rate of $15 per hour. All the experimental procedures were approved by The Institutional Review Board of Columbia University and written informed consent was obtained for each subject.

2.2. Procedure

During the experiments subjects were comfortably seated in a dimly lit room with their head stabilized by a chin-rest at a distance of 54 cm from a computer screen. Eye position and pupil size were measured at a sampling rate of 500 Hz using an Eye-link 1000 eye-tracking system configured for binocular tracking. Before data collection began, the subjects received a task description and performed a few practice trials that were not included in the data set.

In the first part of a session the subjects were required to perform a series of 120 trials in which they read and rated trivia questions and were subsequently shown the answer. The trials were evenly divided between 60 *one-question* trials in which the subjects received a single question, and 60 *two-question* trials, in which they saw two sequentially presented questions and could select the one for which they wished to see the answer. One and two-question trials were signaled in advance by, respectively, one or two "beeps" and were presented in randomly interleaved order in one trial block. A progress bar was displayed after every trial indicating the number of remaining questions.

As shown in Fig. 1 a trial began when the first question was displayed in the upper part of the screen and subjects were asked to rate their levels of curiosity and confidence using a scale of 1 (low) to 5 (high) (panel 1). On 2-question trials, this was followed by the presentation of the second question and its ratings (Fig. 1, panel 2), after which the subjects were prompted to select one question to which they wished to receive the answer using an up/down key press (Fig. 1, panel 3). The trial then progressed to the answer period during which we recorded eye movements as described below (Fig. 1, panels 4–5). After viewing the answer, the subjects received a final rating scale asking them to indicate their surprise in the answer (Fig. 1, panel 6; 1 low, 5 high). One-question trials were identical, except that only one question was displayed and, after giving their curiosity and confidence ratings, the subjects pressed a button to progress to the answer stage.

Our focus was on the subjects' eye movements during a 3 s period centered on answer presentation. To dissociate the anticipatory and reactive components of gaze we divided this period into a 1.5 s *anticipatory epoch* when a rectangular empty box appeared at the top of the screen indicating the forthcoming position of the answer (Fig. 1, panel 4), and a 1.5 s *answer period*, when the answer was displayed aligned to the left edge of the box (Fig. 1, panel 5). All letters (for the questions and answers) were displayed in black with a luminance lower than that of the background, and letter height was approximately 0.39 degrees of visual angle (DVA).

After completing the trivia questions, the subjects completed three questionnaires developed to assess personality traits (110 questions total). The first questionnaire measured the tendency to maximize external or internal sensations on a sensation seeking scale (Zuckerman, 1964). The second questionnaire was the Curiosity and Exploration Inventory II (Kashdan et al., 2009) which measures curiosity and exploration using 2 dimensions: interest for novelty, challenge and absorption (full engagement in specific activities). The third questionnaire analyzes novelty-seeking behaviors on four subscales based on the origin of the stimulation: internal or external to the body, and cognitive versus sensational (Pearson, 1970).

2.3. Data analysis

To study the impact of epistemic curiosity on eye movement patterns we measured eye position as a function of time during the answer period, as well as the number, amplitudes and peak velocities of individual saccades and the number and durations

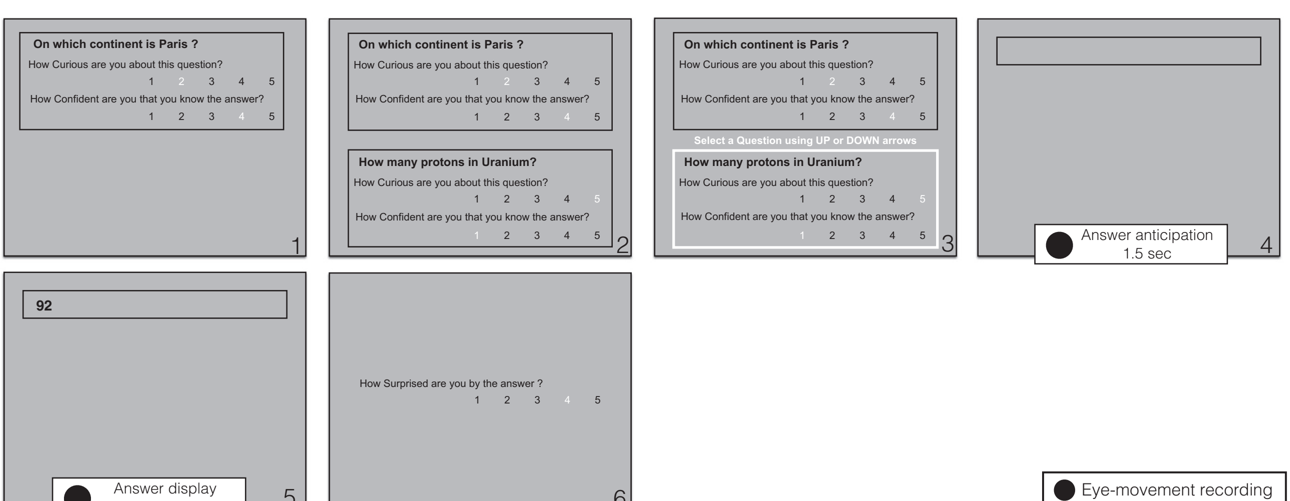

Fig. 1. Task design. The panels illustrate the sequence of events on a 2-question trial: (1) presentation of the first question and curiosity/confidence ratings, (2) presentation of the second question and its curiosity/confidence rating, (3) choice of question, (4) anticipation of answer, (5) presentation of answer, (5) surprise rating.

of individual fixations. Saccades were detected offline as displacements with amplitudes larger than 1 degrees of visual angle and velocity exceeding 0.1 deg/s. A fixation was counted when eye-position was stable for a minimum of 100 ms.

Because subjects had a tendency to emphasize ratings of 1 and 5 for all measures (Fig. 2A–C), we focused on comparing trials with these ratings, which we refer to as, respectively, "low" or "high" curiosity, confidence or surprise.

To analyze the 3 questionnaires, each of which captured a slightly different aspect of curiosity, we summarized the results using a subject-specific *aggregate curiosity score*. To derive this score we first normalized each subject's responses to each questionnaire on a scale from 0 to 100 (i.e., Sensation seeking scale: True = 0, False = 100; Curiosity and exploration inventory: very slightly or not at all = 0; a little = 25; moderately = 50; quite a bit = 75; extremely = 100; Novelty seeking questionnaire: dislike = 0; like = 100), and then computed the average score across the 3 questionnaires.

2.4. Data mining

2.4.1. The random forests algorithm

To determine whether machine learning techniques can learn an accurate predictive model of epistemic states (curiosity, confidence or surprise) we used a common algorithm, the random forest algorithm, which was used in many prior applications. The algorithm showed superior performance in other domains (Criminisi, Shotton & Konukoglu, 2011) including image classification (Bosch, Zisserman & Muoz, 2007), ecology (Cutler et al., 2007) and micro-array analysis (Pang et al., 2006), as well as superior capabilities to identify informative features and disregard redundant ones, model complex predictive interactions between features, and rely on only a few parameters that are easily tuned (Breiman, 2001; Cutler et al., 2007).

The random forests algorithm is based on combinations of classification trees. In one classification tree, each node denotes an individual test on an attribute and branches represent the corresponding outcome, leading progressively to terminal nodes that hold the class labels (e.g., high or low curiosity, confidence or surprise). The algorithm takes as input a vector of eye movement parameters (see below, "*Eye movement parametrization*") and generates several classification trees on a randomly selected subset of features, training them by bootstrapping different versions of the training data. Trees are created incrementally to target training data that are not yet well classified by already constructed trees. Finally, the algorithm fuses the results of all the trees and attempts to classify new data in a test set that had not been included in the training set (Breiman, 2001).

2.4.2. Implementation

We parameterized the Random Forests algorithm using 250 random trees, each using 8 features, and a 10-fold cross-validation method that selects each 10th of data as a test set, while using the remaining data for training. We implemented the

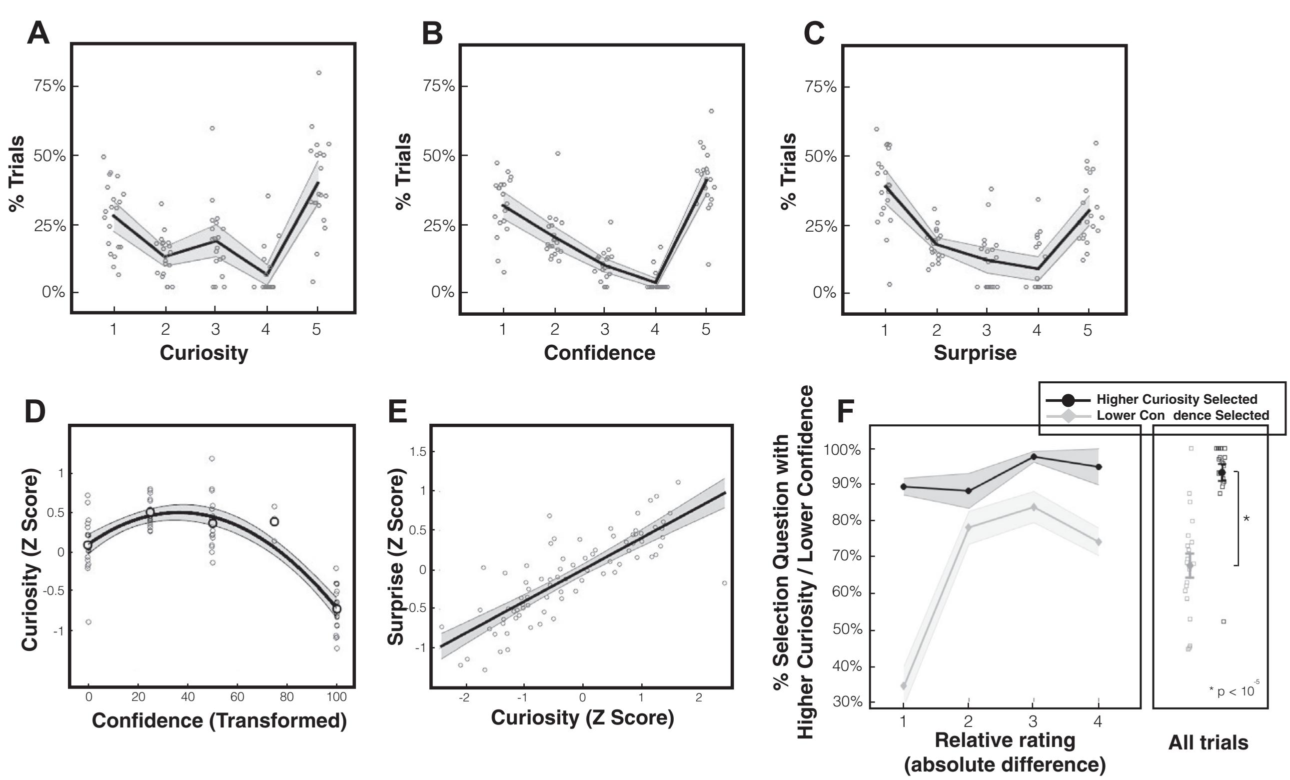

Fig. 2. Ratings in one-question and two-question trials. (A) The distribution of ratings for curiosity (A), confidence (B) or surprise (C). Each point shows the fraction of ratings for one subject across all questions (including 1- and 2 question trials). (D) Curiosity is an inverted-U function of confidence. Curiosity ratings were *z*-scored within each subject, and confidence ratings were transformed to percentages by dividing by 5 and multiplying by 100. Each point represents one trial, and the trials were pooled across one- and 2-question trials and across subjects. The large circles show the average curiosity for each confidence level and the solid and dashed traces show, respectively, the model fit and 95% confidence intervals. The fit produced a significant linear coefficient ($b_1 = 0.021$, 95% confidence interval [0.015, 0.028]) and a nonlinear coefficient as described in the text. (E) Correlation between curiosity and surprise ($r = 0.81$, $p < 10^{-18}$). Same conventions as in panel A. (F) Relative curiosity is a strong predictor of choices on 2-question trials. For each 2-question trial, we calculated relative curiosity/confidence of each pair of questions as the absolute difference between the two ratings. We then plotted the fraction of trials in which subjects chose to see the answer for the question with the higher curiosity (black) or lower confidence rating (gray). Error bars show the SEM across subjects. The two data columns on the right show the values across all levels of relative curiosity/confidence. Each point represents one subject, and the large symbols show the across-subject means and SEM.

analysis in the open-source software "Weka" (Witten & Frank, 2005), which is freely available at http://www.cs.waikato.ac.nz/~ml/weka, and make the full annotated eye movement data set publicly available at: www.gottlieblab.com. Readers can reproduce our analysis by using Weka with parameters –I 250 –K 8 –S 1.

We conducted the mining analyses separately for predicting ratings of curiosity, confidence and surprise. In a first *population analysis* we used all trials with ratings of 1 or 5, pooled across all observers. Because any imbalance in the frequency of high and low ratings can artificially inflate prediction accuracy, we resampled the pooled data set to provide equal number of trials with each type of rating, thereby ensuring that chance accuracy was 50%. After resampling we were left with 990 trials for the analysis of curiosity, 1144 trials for confidence and 1042 trials for surprise. In this remaining data set, the number of trials per subject was too low to provide a meaningful evaluation of how accurate a machine learning algorithm is for individual subjects. Therefore, to verify the consistency of the algorithm, we used a standard *cross-subject validation procedure*, whereby we trained the algorithm on 19 of the 20 subjects and tested its prediction accuracy on the remaining subject, iterating until each subject served as a test. To assess statistical significance we tested the set of predictions against 0.5 using non-parametric tests. Note that, although the training sets overlap across the 20 subjects, the results of the test phase are based on the individual subject data, and therefore produce 20 independent points, satisfying the independence assumption for statistical tests.

To determine the prediction value of individual eye movement features we used the "GainRatio" tool provided by Weka. The tool outputs the information gain ratio (IGR), defined as the reduction in entropy achieved by using information only from that feature:

$$\begin{aligned} \mathrm{IGR} &= \mathrm{GainR}(\mathrm{Class}, \mathrm{feature}) \\ &= (\mathrm{H}(\mathrm{Class}) - \mathrm{H}(\mathrm{Class}|\mathrm{Feature}))/\mathrm{H}(\mathrm{Feature}) \end{aligned}$$

where "class" refers to the categories to be predicted (high or low curiosity, confidence or surprise) and H() denotes the entropy of the respective distribution (see Witten & Frank, 2005).

2.4.3. Eye movement features

While traditional data analyses can only focus on a small number of features, data mining techniques based on machine learning can rapidly evaluate very large sets of features and feature combinations. Therefore, these techniques can reveal new parametrization of the eye movement traces that make efficient statistical predictions but may not have been identified on an *a priori* basis.

We parametrized the input to the algorithm using a vector of 183 features that described the statistical properties of the eye movement traces on each trial. A complete list of features is given in Supplementary Table 1 and a general description is given below.

The majority of the features were statistical properties of 4 signals: (1) the horizontal position of the eye in screen coordinates, (2) the vertical position of the eyes in screen coordinates, (3) pupil diameter (*z*-scored for each subject by subtracting the mean and dividing by the standard deviation across all the trials given by that subject), and (4) the Euclidean distance between the eye and the upper left corner of the answer box. Each signal was measured during the 1.5 s before-answer and after-answer periods at a frequency of 500 Hz, providing 2 data sets of 750 data points each.

For each of the 8 resulting data sets (4 signals * two epochs), we extracted a set of 40 basic features before the answer and after answer onset (listed, respectively, in lines 1–40 and 84–123 in Supplementary Table 1). These values were based on statistical properties of the measurements, including the minimum, maximum, mean, median, variance, range (max–min), and the first quartile, third quartile and interquartile range (difference between the first and second quartiles). In addition, to capture aspects of gaze dynamics during the two epochs, we computed a vector of local derivatives (differences between successive time points) and extracted the mean of the absolute value of the derivatives in each epoch. Finally, we provided the difference between the "before answer" and "after answer" values for the 20 features characterizing the distance to answer and the pupil size (features 164 to 183). For measures of pupil size, to mitigate noise induced by variations in gaze location or luminance, we extracted additional descriptors from a subset of trials when the eyes were continuously inside the answer box during the interval 0.5 s before to 0.5 s after answer onset. The 10 features extracted from the first and second halves of this interval (250 samples each) are listed in lines 74 to 83 (before answer), and 154 to 174 (after answer).

Finally, we extracted features describing individual saccades and fixation: 6 features for each of the 5 first fixation of epochs before the answer (Supplementary Table 1, lines 41 to 70) and after the answer display (lines 124 to 153). For each fixation we extracted (1) the time of beginning of the fixation (the end of the pre-fixation saccade), (2) the end of each fixation (the onset of the post-fixation saccade), (3) fixation duration (difference between start and end times), (4) the average horizontal position during the fixation period in screen coordinates, (5) the average vertical position during the fixation period in screen coordinates, (6) the average Euclidean distance between the fixation location and the answer box. If a trial included fewer than 5 fixations in either epoch, the missing numbers were coded as NaN (not a number). Finally, we computed the time when the eyes first entered the answer box (71 if this time was before, and line 72 if it was after answer onset) and the duration of the first fixation in this area (line 73).

3. Results

3.1. Ratings and choices

As shown in Fig. 2A–C, the distribution of ratings was not uniform but tended to be concentrated on values of 1 and 5 for curiosity, confidence and surprise. This suggests that subjects were engaged in the task and did not merely settle on a default intermediate rating. Because of this asymmetry, we focused our eye movement analyses on trials with ratings of 1 or 5 (which we refer to as "low" or "high").

Previous investigations proposed that a key factor driving epistemic curiosity is an "*information gap*" – a discrepancy between what one knows and what one would like to know – and therefore that curiosity should peak when one has a little bit of knowledge, but diminish if a subject knows too little or too much about a topic (Kang et al., 2009; Lowenstein, 1994). Consistent with this prediction, we found that ratings of curiosity were an inverted U-shaped function of confidence – peaking at intermediate levels but becoming lower for the lowest or the highest levels of confidence (Fig. 2D). Following the method of Kang et al. (Kang et al., 2009), we transformed confidence ratings to percentages (dividing each rating by 5 and multiplying by 100), and fit the data to the equation: curiosity = $b_0 + b_1 * p + b_2 * p * (1 - p)$, where p is the transformed confidence score. The model provided $R^2 = 0.73$ and a significant quadratic coefficient ($b_2 = 30 * 10^{-4}$, 95% confidence interval, $[24 * 10^{-4}, 36 * 10^{-4}]$) consistent with a peak at intermediate values. We obtained equivalent results whether we used the raw or *z*-scored confidence ratings.

Ratings of curiosity were positively correlated with ratings of surprise (Fig. 2E; $r = 0.982$, $p = 0.003$). Interestingly, subjects could have understood their ratings of "surprise" in this task to indicate how unlikely they believed the answer to be (how much it differed from the set of possible answers) or alternatively, how much weight they gave to the new information. The high correlation between curiosity and surprise suggests that subjects adopted the latter interpretation – i.e., the more curious they were while anticipating the answer, the more salience they assigned to the answer when it was finally shown. In the following analyses we show that, despite this high correlation between curiosity and surprise, the two constructs have different oculomotor effects.

To investigate how the subjects' ratings were related to their choices on 2-question trials, we calculated for each trial the absolute difference between the ratings of curiosity and confidence assigned to the two alternative questions. We reasoned that, if the relative ratings determined the subjects' choices, subjects may tend to choose the question with the higher curiosity or lower confidence on each trial.

As shown in Fig. 2F, this hypothesis was confirmed and choices were determined more strongly by ratings of curiosity than by those of confidence. Subjects almost always asked for the answer to the question that produced the higher curiosity, even if the rating differences were small (Fig. 2F; black trace). While the subjects also tended to select the question in which they had lower confidence, this was a weaker effect (Fig. 2F, gray). If both questions elicited a similar level of confidence (a difference of 1), choices were fully allocated to the question that elicited higher curiosity. Across the entire set of 2-question trials (Fig. 2F, right panel), subjects selected the question with the higher curiosity rating on 93% ± 2.3% of trials and the question with the lower confidence rating on only 68% ± 3.2% of trials ($p < 10^{-5}$, Wilcoxon test). Together, these findings suggest that confidence influences curiosity according to an inverted-U function (Fig. 2D) but it is curiosity that ultimately determines the subjects' choice of information.

3.2. Eye movements

To examine the overall eye movement patterns during the task, we constructed time-resolved maps of fixation density at each screen location (Fig. 3). Consistent with previous investigations, subjects allocated their gaze behavior in a task-related fashion, both anticipating and reacting to the task events (Tatler et al., 2011). At the onset of the anticipation period, gaze position was diffusely distributed in the upper half of the screen – where the subjects had just viewed the question and provided their ratings – whereas thereafter gaze became increasingly clustered on the left corner of the answer box – where the answer was expected to be. Gaze transiently converged on the answer after it appeared, and then gradually drifted toward the center of the screen – the

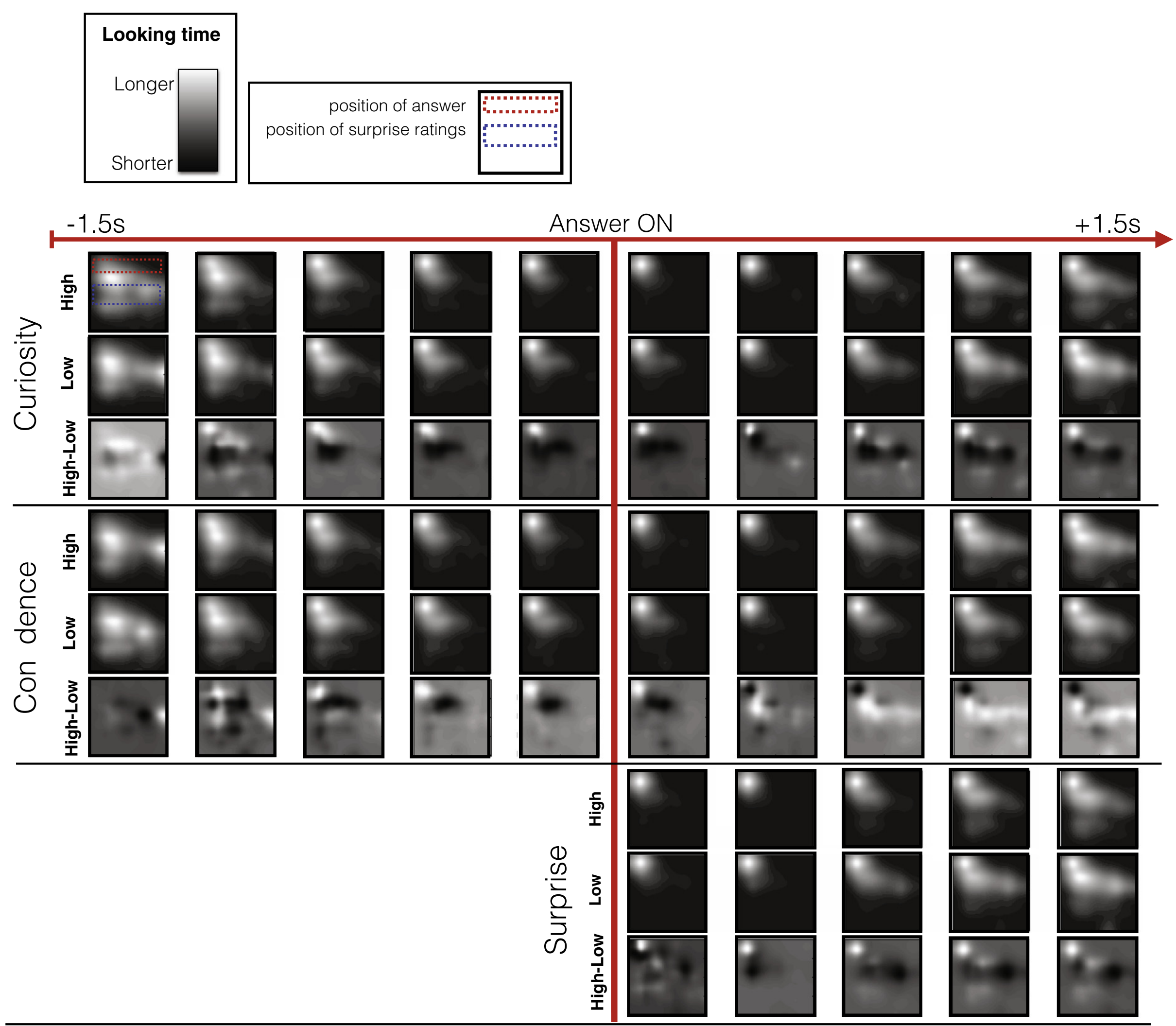

Fig. 3. Probability density of fixations during the answer epoch. We divided the 3 s of eye position recording in 10 non-overlapping 300 ms bins, and divided the total area of the screen display into pixels measuring 0.27 by 0.15 degrees of visual angle. We then calculated the fraction of time, out of each time bin, that a pixel was fixated. Values were averaged for each subject and then averaged across subjects. The heat maps show group averages of trials with ratings of 1 or 5. The subtraction maps subtract two individual maps such that brighter areas correspond to zones that receive more exploration for the higher rating.

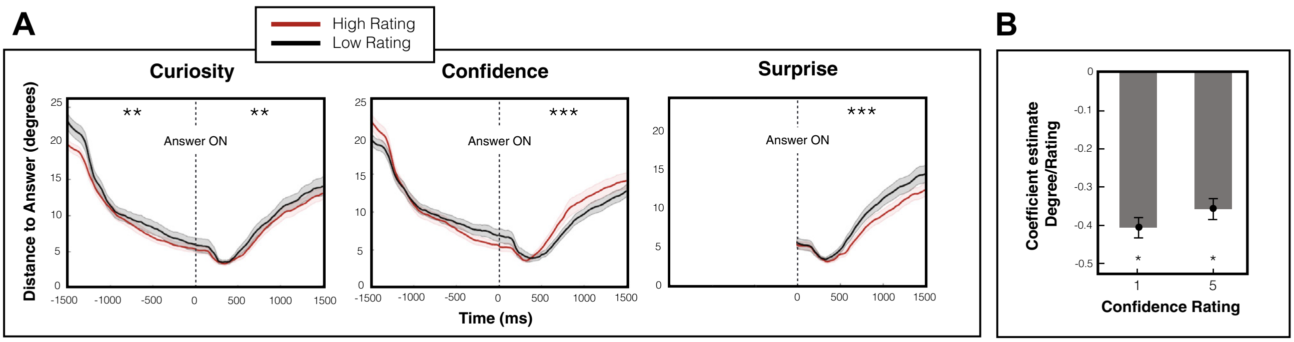

Fig. 4. Distance to answer. (A) For each trial with high or low ratings we computed the distance between the eye position and the left edge of the answer box every 2 ms. Distances were averaged for each subject, and we display the mean and SEM across subjects. Average distances before and after answer onset were compared with a 1-way ANOVA; stars show $^{**}p < 10^{-45}$, $^{***}p < 10^{-75}$. (B) To examine whether the effect of curiosity was robust to variations in confidence, we divided trials into those with high and low confidence ratings and computed a regression analysis to determine the impact of curiosity in each group (using average eye position in the 500 ms before answer onset). The panel shows the regression coefficients and their 95% confidence interval, and the stars indicate $p < 10^{-30}$. Negative coefficients indicate that higher curiosity ratings were associated with smaller distance to the answer box.

Table 1
Comparison of saccade and fixation metrics as a function of epistemic states.

		Curiosity		Confidence		Surprise	
		High	Low	High	Low	High	Low
Anticipatory	#Saccades	3.31(0.17)	3.33(0.13)	3.28(0.12)	3.74(0.18)	3.47(0.15)	3.34(0.13)
	Sac. Amp. (deg.)	8.25(0.63)	9.00(0.62)	9.11(0.58)	7.78(0.49)	8.43(0.53)	8.92(0.55)
	Peak. Vel. (deg/s)	240.78(11.83)	249.21(9.54)	243.56(8.27)	235.05(11.20)	245.36(11.12)	252.04(10.09)
	#Fix.	2.92(0.12)	2.90(0.09)	2.89(0.11)	3.25(0.13)	3.03(0.10)	2.93(0.09)
	Fix. Dur. (ms)	0.36(0.02)	0.35(0.02)	0.36(0.01)	0.33(0.02)	0.34(0.02)	0.35(0.02)
After answer	#Saccades	3.51(0.13)	3.55(0.13)	3.49(0.11)	3.48(0.12)	3.45(0.16)	3.56(0.13)
	Sac. Amp. (deg)	8.39(0.53)	7.60(0.54)	7.54(0.56)	7.42(0.43)	8.03(0.52)	7.77(0.58)
	Peak. Vel. (deg)	240.54(9.67)	234.95(12.07)	228.94(11.18)	228.05(9.68)	246.66(12.21)	228.60(11.39)
	# Fix.	2.82(0.13)	2.88(0.11)	2.76(0.09)	2.82(0.09)	2.69(0.12)	2.81(0.10)
	Fix. Dur. (ms)	0.28(0.02)	0.30(0.02)	0.30(0.01)	0.29(0.01)	0.27(0.02)	0.30(0.02)

Each entry shows the average (SEM) of the respective metric in 1-question trials, first averaged by subject and then across subjects. For each metric we compared the high and low rating trials using a Wilcoxon test. While higher confidence tended to be associated with fewer saccades, fewer fixations and larger saccade amplitudes during the anticipation period, these trends did not reach statistical significance (p-values of, respectively, 0.063, 0.07 and 0.08). For all other comparisons, $p > 0.13$.

anticipated location of the surprise ratings (whose presentation is not included in Fig. 3).

In addition to this task-related pattern, gaze was modulated by curiosity, confidence and surprise. This can be appreciated by comparing the top two rows of fixation maps in each category, corresponding to "high" and "low" ratings, and from the subtraction maps on the 3rd row. These displays suggest that gaze converged sooner and lingered longer on the expected answer location for high relative to low curiosity states.

To quantitatively analyze these effects and distinguish them from those of confidence and surprise, we measured the average Euclidean distance between the eye and the top left corner of the answer box as a function of time (Fig. 4A). This distance was significantly smaller for questions with high relative to low curiosity ratings, and this difference was maintained consistently during the entire anticipation epoch ($p < 10^{-45}$, Wilcoxon test). Therefore, subjects shifted gaze sooner to the answer box if they had higher curiosity. In contrast, confidence had an inconsistent effect that switched from early repulsion to a slight later attraction in high versus low confidence states, but was not significant over the entire interval (Fig. 4A, middle). To further confirm that the effects of curiosity are not confounded by those of confidence, we carried out a second analysis where we computed the effects of curiosity separately in trials with high and low confidence ratings. Linear regression coefficients computed over the 500 ms before answer onset were significantly negative in both trial subgroups ($p < 10^{-30}$ in both cases), showing that high curiosity was associated with a shorter distance to the answer independently of confidence ratings (Fig. 4B).

Eye movements were also influenced by subjective ratings after answer presentation – lingering longer on the answer for trials with higher curiosity, lower confidence and higher surprise (Fig. 4A). However, because of the statistical associations among these ratings (Fig. 2D and E), these post-answer effects cannot be unambiguously attributed to a specific factor – a topic to which we return in the data mining analyses below.

As shown in Table 1, curiosity, confidence and surprise had no significant effects on other gaze parameters, including the number, amplitudes, peak velocities of saccades, or the number and durations of fixations, in the epochs that preceded or followed the answer onset. Therefore in higher curiosity states, subjects seem to have guided their gaze more precisely toward the answer box without altering the speed or frequency of saccades.

3.3. Correlation with trait curiosity

To determine whether the impact of curiosity on eye movements correlates with measures of trait curiosity (Risko et al., 2012), we constructed an aggregate curiosity score based on the subjects' answers to 3 questionnaires designed to measure sensation seeking, curiosity and exploration and novelty seeking traits (see Section 2). We found a negative correlation between the questionnaire score and the effect of curiosity on saccadic anticipation (Fig. 5A; linear regression coefficient, −0.079; SE = 0.031, $p = 0.02$). By contrast, there were no significant correlations between questionnaire scores and the extent to which gaze was affected by confidence (Fig. 5B; linear regression coefficient, 0.032; SE = 0.246, $p = 0.17$) or surprise after answer onset

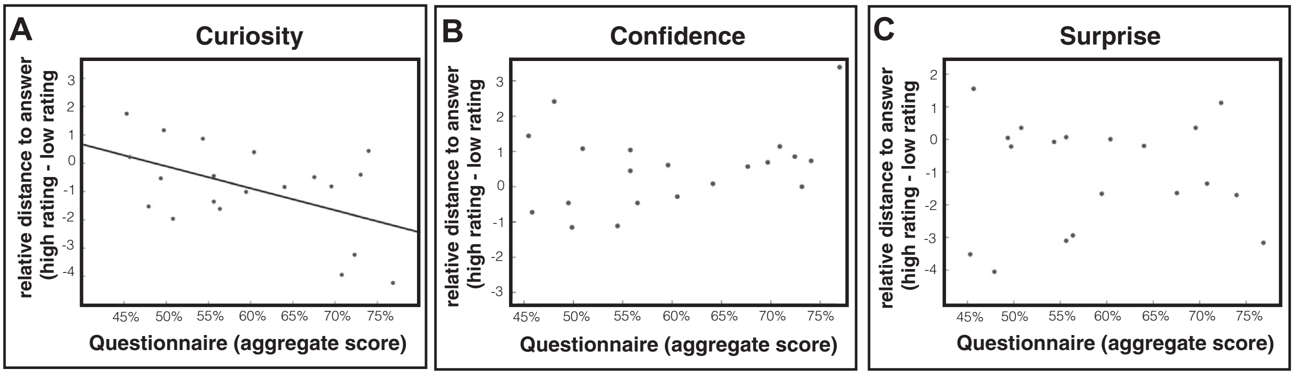

Fig. 5. Effects of state curiosity on eye movements correlate with trait curiosity. To measure the correlation between questionnaire ratings and oculomotor effects we computed for each subject the average eye distance to answer (over the 3-s of eye-movement recording for curiosity and confidence, and over 1.5 s after answer onset for surprise). We then computed an effect size for each subject as the average difference between trials with ratings of 5 versus 1, and plotted it against the subject's aggregate questionnaire score (see Section 2). Linear regression (black line) was significant for the first panel but not for the other two (omitted).

(Fig. 5C; −0.027; SE = 0.042, $p = 0.53$). Therefore trait curiosity correlates with the impact of curiosity on gaze: subjects showing higher trait curiosity as measured by questionnaires also showed a stronger tendency to anticipate the answer in high versus low curiosity states.

3.4. Data mining

While the analyses we conducted so far show that ratings of curiosity, confidence and surprise affect eye movement patterns, in the following section we asked the converse question: are the eye movements effects specific enough to provide a read out of the subject's epistemic state? Please note that these two questions are mathematically distinct and are expected to provide non-redundant information. Traditional analyses are equivalent to estimating the conditional probability of an *eye movement pattern given an epistemic state* (e.g., *P*(eye movement | high or low rating)), whereas a data mining approach estimates the inverse probability – the likelihood of *an epistemic state given an eye movement pattern* (e.g., *P*(high or low rating | eye movement)). In addition, while a traditional approach focuses on a small set of "intuitive" features that are believed to be relevant based on prior hypotheses, data mining algorithms sift through very large sets of features and feature combinations, and can reveal new parameterizations of the input (eye movement) record that are not necessarily intuitive but support efficient predictions.

Table 2
Performance parameters for the random forests algorithm.

	Class	TP rate	FP rate	Precision	Recall	*F*-measure
Curiosity	*% correctly classified instances: 69.8%; ROC area: 0.753*					
	Low	0.689	0.293	0.702	0.689	0.0695
	High	0.707	0.311	0.694	0.707	0.701
	Average	0.698	0.302	0.698	0.698	0.698
Confidence	*% correctly classified instances: 72.8%, ROC area: 0.796*					
	Low	0.771	0.315	0.71	0.771	0.739
	High	0.685	0.229	0.75	0.685	0.716
	Average	0.728	0.272	0.73	0.728	0.728
Surprise	*% correctly classified instances: 63.1%; ROC area 0.687*					
	Low	0.614	0.353	0.635	0.614	0.624
	High	0.647	0.386	0.626	0.647	0.636
	Average	0.631	0.369	0.631	0.631	0.63

TP rate: rate of true positives (fraction correctly classified as a given class); FP rate: rate of false positives (fraction falsely classified as a given class); Precision: proportion of instances that are truly of a class divided by the total instances classified as that class; Recall: proportion of instances classified as a given class divided by the actual total in that class (equivalent to TP rate); *F*-measure: a combined measure for precision and recall calculated as 2 * Precision * Recall/(Precision + Recall).

In an initial application of the random forest data mining algorithm we derived predictions based on the entire eye movement

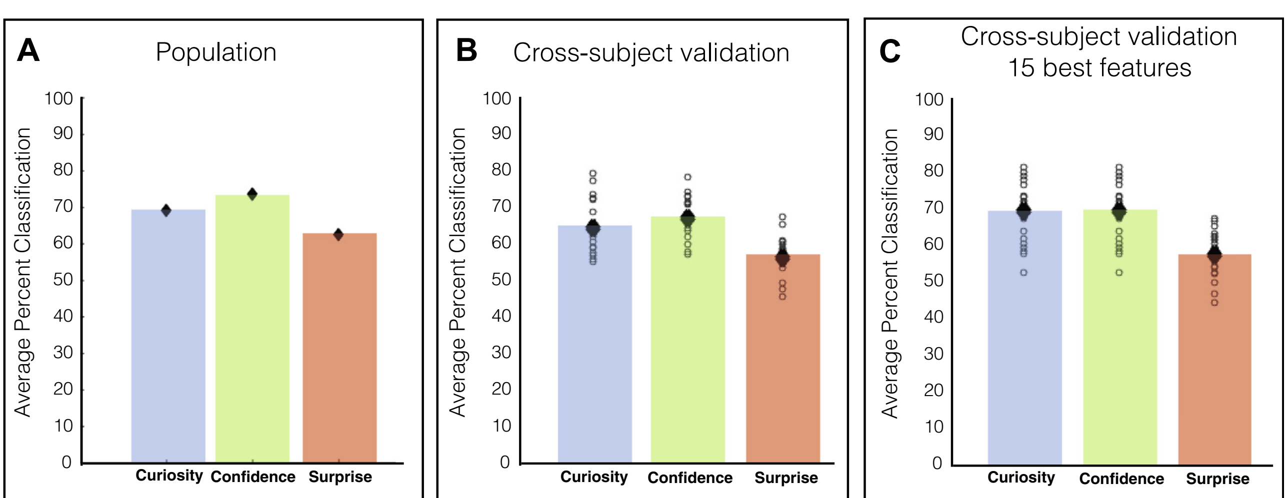

Fig. 6. Classification accuracy for different implementations. (A) Classification across the entire data set. (B) Classification with across-subject cross-validation. (C) Same as B but using only the 15 features with the highest IGR. In (B) and (C), the open points show individual subject predictions and the black points and bars show average and SEM.

data set, pooled across observers. This produced above-chance classification accuracy of 69.79%, 72.81% and 63.05% for, respectively, curiosity, confidence and surprise (Fig. 6A; recall that chance levels were 50% in the re-sampled data set; see Section 2). Details of the algorithm performance are shown in Table 2, and establish that performance was unbiased and of similar accuracy for low and high rating trials. To examine the reliability of the algorithm across individual observers we used a cross-subject validation procedure, whereby predictions were derived for each individual subject based on training on the remaining subjects (see Section 2). Individual prediction accuracy was similar to that obtained across the population and, across subjects, was significantly higher than chance (Fig. 6B, mean and SEM were, for curiosity: 0.65 ± 0.02, $p < 10^{-6}$ relative to 0.5 (Wilcoxon test); confidence 0.68 ± 0.01, $p < 10^{-9}$; and surprise 0.57 ± 0.01, $p < 10^{-4}$). Finally, the cross-validation procedure produced similar results when it was replicated using only the 15 most informative features (see below; Fig. 6C), showing that the algorithm achieves above-chance performance based on a small number of predictive features that generalize across multiple subjects.

To investigate the basis of the classification performance we examined the gain in prediction – the IGR metric – afforded by individual features (see Section 2). Of the set of 183 features that the algorithm received as input, only a small subset was identified as having high IGR (Fig. 7A). Replotting the IGR values in the order of features showed that these features clustered in several groups (Fig. 7B). The most informative features for predicting curiosity, confidence or surprise were related to the eye position, primarily before the answer (features 1–30) and, to a lesser extent also after answer onset (features numbers 85–113; see in Supplementary Table 1). Measures of pupil dilation were only weakly informative after answer onset (feature 114–123), consistent with the fact that pupil diameter covaried weakly with the curiosity ratings (Supplementary Fig. 1).

To compare these results with those that would be expected from a traditional analysis, we computed the difference in average

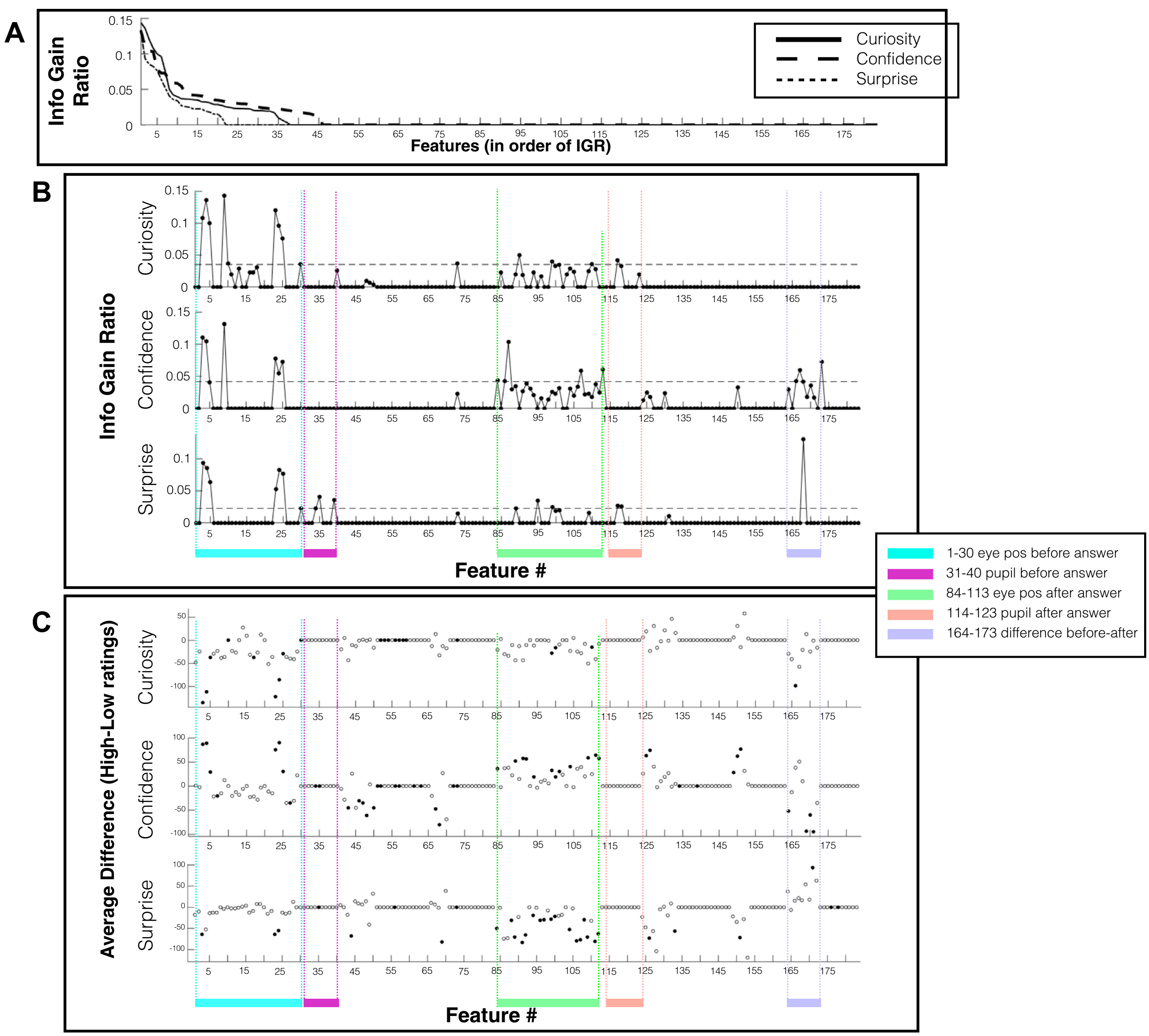

Fig. 7. Information gain ratios from the random forests algorithm. (A) The IGR values produced by the algorithm sorted in order of magnitude for each of the 3 ratings. (B) The same set of IGR values sorted as a function of the feature number. The dotted horizontal lines show the threshold separating the 15 features with the highest IGR from the remaining features. Color bars denote selected feature categories as detailed in Supplementary Table 1. (C) Effect of rating on the 183 features. For each of the features provided to the data mining algorithm, we computed the difference between the average values on trials with high minus those with low ratings. Filled circles show differences that were significant at $p < 0.05$. Note that different features spanned different ranges of numerical values; in this plot, some significant differences appear to be 0 because they were very small relative to the ordinate scale (which was chosen so as to incorporate the numerical range spanned by all the features). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
The 15 most informative features for the random forests algorithm.

Curiosity			Confidence			Surprise		
IGR	Feature #	Feature ID	IGR	Feature #	Feature ID	IGR	Feature #	Feature ID
0.1438	9	bXinterQuartileRange(1)	0.1314	9	bXinterQuartileRange(1)	0.1301	168	cDV(1)
0.1359	4	bXRange(1)	0.1105	3	bXMax(1)	0.0933	3	bXMax(1)
0.1206	23	bDMax(1)	0.1041	4	bXRange(1)	0.0854	4	bXRange(1)
0.1081	3	bXMax(1)	0.1032	87	aXRange(1)	0.0825	24	bDRange(1)
0.0997	5	bXV(1)	0.0776	23	bDMax(1)	0.0764	25	bDV(1)
0.0962	24	bDRange(1)	0.073	25	bDV(1)	0.0639	5	bXV(1)
0.0755	25	bDV(1)	0.0726	173	cDMDerv(1)	0.052	23	bDMax(1)
0.0500	90	axFirstQ(1)	0.0608	113	aDMDerv(1)	0.0408	35	bPSV(1)
0.0418	117	aPSRange(1)	0.0591	167	cDRange(1)	0.0354	39	bPSinterQuartileRange(1)
0.0397	99	aYMedian(1)	0.0584	107	aDRange(1)	0.0341	95	aYMin(1)
0.0368	10	bXMDerv(1)	0.0544	24	bDRange(1)	0.0269	117	aPSRange(1)
0.0363	73	aFDur(1)	0.0438	84	aXMean(1)	0.0258	118	aPSV(1)
0.0361	110	aDFirstQ(1)	0.0421	86	aXMax(1)	0.0246	99	aYMedian(1)
0.0355	30	bDMDerv(1)	0.0421	166	cDMax(1)	0.0225	30	bDMDerv(1)
0.0350	101	aYThirdQ(1)	0.0415	168	cDV(1)	0.0225	89	aXMedian(1)

The feature # and feature ID refers to the entries in Supplementary Table 1.

values between high and low rating trials for each of the 183 features used for data mining (Fig. 7C). Comparison of Fig. 7B and C highlights several noteworthy points. First, while many features are modulated as a function of ratings (show differences above or below 0 in Fig. 7C) only some of these features have high IGR (Fig. 7B). The features with high IGR are those that modulate in a manner that is specific enough to allow accurate predictions. Notably, features describing eye position *before* the answer onset (features 1–30) tend to modulate in distinct manners for the 3 ratings (e.g., change in opposite directions for curiosity and confidence and be unaffected by surprise) and have high IGR. In contrast, features describing eye position after the answer (features 84–113) show correlated variability for curiosity and surprise and thus have lower IGR. This result illustrates how data mining techniques can help interpret empirically measured effects, and upholds our conclusion based on Fig. 4 that the anticipatory, rather than reactive, component of gaze can most reliably distinguish between epistemic states.

Second, for the anticipatory component of gaze (features 1–30), data mining assigns high IGR to features related to the *variability* rather than the central tendencies of the eye position – including the interquartile range, the absolute range, the maximum value and the standard deviation of the horizontal eye positions and the distance to answer (Table 3). Examination of Fig. 4 shows that this finding extends rather than conflicting with our initial analysis. While in our initial analysis (Figs. 3 and 4) we focused on the average distance to answer, the results can be equally well described by variance statistics: across the anticipatory epoch, the range of distances between the eye and the answer box was larger for low relative to high curiosity states, due primarily to the maximum distance (at −1500 ms) being larger for the former set. Indeed, the vast majority of the variance-related features that had high IGR also showed significant differences between high and low curiosity trials when re-examined in Fig. 7C.

In sum, as expected, the data mining analysis upholds our conclusion that curiosity states affect the anticipatory component of gaze, and identifies new parameterizations of the eye movement trace that support efficient statistical classification.

4. Discussion

In addition to its documented ability to enhance memory and motivation (Gruber et al., 2014; Kang et al., 2009), epistemic curiosity influences attention and gaze. On some level this finding seems unsurprising – as it seems natural to look more intently at information that we are more curious about. However, no studies have quantitatively examined the link between eye movements and semantic curiosity, and recent studies have remained agnostic about (Gruber et al., 2014) or even argued *against* such a link (Kang et al., 2009). Our results provide empirical data showing that semantic curiosity affects eye movement control and that these effects are specific enough to allow curiosity to be read out independently of other epistemic variables such as confidence and surprise. This finding, combined with a previous report linking trait curiosity with free-viewing visual exploration (Risko et al., 2012), suggests that curiosity has multiple influences on eye movement control.

An important question concerns the relation between curiosity and surprise, because novelty and surprise have been shown to impact eye movement control (Baldi & Itti, 2010; Itti & Baldi, 2009; Yang, Chen, & Zelinsky, 2009), and in our results these two ratings were significantly correlated (Fig. 2B). However, whereas prior studies focused on visual paradigms, here we focus on semantic factors. Most relevant to our work, two papers focused on Bayesian surprise have defined surprise in a purely bottom-up fashion – based on the conditional probability of observing a pixel given its local visual context (Baldi & Itti, 2010; Itti & Baldi, 2009). By this definition, surprise is very closely related to visual salience (contrast) – and is clearly distinct from the epistemic surprise we examine here. Second, whereas previous investigations focused on the reactive component of gaze – the propensity to look at a novel or surprising item – our paradigm revealed a *pro-active* response, whereby curiosity enhanced gaze anticipation *before* the subjects saw the answer and could rate its "surprise". This anticipatory component – which preceded any effect of surprise – was the basis for dissociating the effects of curiosity from those of surprise both in the traditional analyses (Figs. 3 and 4) and in the data mining approach (Fig. 7).

Based on these considerations, we suggest that curiosity can be viewed as a pro-active process that anticipates, or motivates agents to obtain new information, whereas surprise indicates a reactive process after having processed the information. Together with the information gap theory that suggests that curiosity peaks at intermediate levels of knowledge (see Fig. 2A and (Kang et al., 2009; Lowenstein, 1994)), this reinforces the view of curiosity as a mechanism for active learning, which allows agents to proactively choose which questions they wish to resolve, and specifically seek out learnable tasks while steering away from unlearnable or boring information (Gottlieb et al., 2013; Oudeyer et al., 2013).

The fact that curiosity influences eye movements is particularly important because it provides a possible handle into its cellular mechanisms. In humans and non-human primates, oculomotor decisions are thought to be mediated by "priority maps" – populations of neurons in the lateral intraparietal area and the frontal eye fields that have visuospatial receptive fields and select targets for attention or gaze, which are sensitive to multiple factors, including bottom-up salience, task relevance and expected rewards (Bisley & Goldberg, 2010; Thompson & Bichot, 2005). Our results suggest that epistemic curiosity should be added to the list of factors that determines priority and attention allocation. The finding that epistemic curiosity specifically affects gaze anticipation supports the idea that it acts at the oculomotor decision stage rather than lower levels of motor control. Thus, the mechanisms that generate curiosity, including its individual variations, may be read out in neurons involved in cognitive eye movement control.

Our demonstration that machine learning algorithms can read out states of curiosity from eye movement patterns reinforces the conclusion that curiosity produces an oculomotor signature that is distinct from that of partially correlated constructs of confidence and surprise. Of the large number of eye movement features provided to algorithm, only a small subset was predictive of epistemic states, and this subset supported efficient classification across individual observers (Fig. 6C), suggesting that the algorithm is robust to changes in the precise features it receives (e.g., would produce similar results if given a slightly different set of eye movement features). This finding adds to the growing body of research that applies data mining techniques to emotion recognition (e.g., based on facial expressions (Zeng, Pantic, Roisman, & Huang, 2009), speech intonation (Oudeyer, 2003) or skin conductance (Jerritta, Murugappan, Nagarajan, & Wan, 2011)). The ability to predict curiosity states has potential practical applications in several domains, including the development of individualized education software (Clement, Roy, Oudeyer, & Lopes, 2014) where online readouts of curiosity could be used to customize the content of instruction that is shown to the learner so as to maximize his/her individual learning progress, and of recommender systems (Rokach, Shapira, & Kantor, 2011) where tracking the curiosity of customers may be used to offer personalized product recommendations. Finally, curiosity is strongly related to early infant learning and development (Oudeyer & Smith, in press) and the tracking of curiosity could become an important tool in efforts to leverage non-invasive techniques in research and diagnosis of developmental disorders such as attention deficit disorders and autism (Gliga, Bedford, Charman, Johnson, & Team, in press).

Disclosure statement

The authors declare that they have no conflict of interest.

Acknowledgments

This work was funded by a Fulbright visiting scholar grant (AB), HSFP Cross-Disciplinary Fellowship LT000250 (AB), and Inria Neurocuriosity grant (AB, PYO, JG). Special thanks to Drs. Hakwan Lau and D. Graham Burnett for the use of the Eyelink eye tracker. We thank Latoya Palmer and Cherise Washington for expert administrative assistance, and to Manuel Lopes and the members of the Mahoney Center who read and had helpful suggestions on the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2015.10.009.

References

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks, 23*(5), 649–666.
Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology, General Section, 45*(3), 180–191.
Bisley, J., & Goldberg, M. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33*, 1–21.
Bosch, A., Zisserman, A., & Muoz, X. (2007). Image classification using random forests and ferns. In *IEEE 11th international conference on computer vision* (pp. 1–8).
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
Cardoso-Leite, P., & Bavelier, D. (2014). Video game play, attention, and learning: How to shape the development of attention and influence learning? *Current Opinion in Neurology, 27*(2), 185–191.
Clement, B., Roy, D., Oudeyer, P.Y., & Lopes, M. (2014). Online optimization of teaching sequences with multi-armed bandits. *7th International Conference on Educational Data Mining*.
Cutler, D. R., Edwards, T. C., Jr, Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783–2792.
Gliga, T., Bedford, R., Charman, T., Johnson, M., & Team, T.B. (in press). Enhanced visual search in infancy predicts emerging autism symptoms. *Current Biology*.
Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron, 76*(2), 281–295.
Gottlieb, J., Oudeyer, P.Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and empirical mechanisms. *Trends in Cognitive Science*, in press.
Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, reward and information seeking. *Journal of Neuroscience, 34*(46), 15497–154504.
Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84*(2), 486–496.
Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295–1306.
Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011). Physiological signals based human emotion recognition: A review. *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium* (pp. 410–415).
Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., et al. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973.
Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. T., Breen, D. T., & Steger, M. F. (2009). The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. *Journal of Research on Personality, 43*(6), 987–998.
Lowenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75–98.
Oudeyer, P.-Y. (2003). The production and recognition of emotions in speech: Features and algorithms. *International Journal of Human-Computer Studies, 59*(1), 157–183.
Oudeyer, P.Y., & Smith, L. (in press). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*.
Oudeyer, P.-Y., Baranes, A., & Kaplan, F. (2013). Intrinsically motivated learning of real-world sensorimotor skills with developmental constraints. In *Intrinsically motivated learning in natural and artificial systems* (pp. 303–365). Springer.
Pang, H., Lin, A., Holford, M., Enerson, B. E., Lu, B., Lawton, M. P., & Zhao, H. (2006). Pathway analysis using random forests classification and regression. *Bioinformatics, 22*(16), 2028–2036.
Pearson, P. H. (1970). Relationships between global and specified measures of novelty seeking. *Journal of Consulting and Clinical Psychology, 34*(2), 199–204.
Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition, 122*, 86–90.
Rokach, L., Shapira, B., & Kantor, P. B. (2011). *Recommender systems handbook* (Vol. 1). New York: Springer.
Tatler, B. W., Hayhoe, M. N., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5), 5–25.
Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research, 147*, 251–262.
Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
Yang, H., Chen, X., & Zelinsky, G. J. (2009). A new look at novelty effects: Guiding search away from old distractors. *Attention Perception and Psychophysics, 71*(3), 554–564.
Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual and spontaneous expressions. *IEEE Transactions, 7th International Colloquium on Patterns Analysis and Machine Intelligence, 31*(1), 39–58.
Zuckerman, M. (1964). Development of a sensation-seeking scale. *Journal of Consulting and Clinical Psychology, 28*(6), 477.