Defending Yarbus: Eye movements reveal observers' task

Ali Borji

Department of Computer Science,
University of Southern California, Los Angeles, CA, USA

Department of Computer Science,
University of Southern California, Los Angeles, CA, USA
Neuroscience Graduate Program,
University of Southern California, Los Angeles, CA, USA
Department of Psychology,

Laurent Itti

University of Southern California, Los Angeles, CA, USA

In a very influential yet anecdotal illustration, Yarbus suggested that human eye-movement patterns are modulated top down by different task demands. While the hypothesis that it is possible to decode the observer's task from eye movements has received some support (e.g., Henderson, Shinkareva, Wang, Luke, & Olejarczyk, 2013; Iqbal & Bailey, 2004), Greene, Liu, and Wolfe (2012) argued against it by reporting a failure. In this study, we perform a more systematic investigation of this problem, probing a larger number of experimental factors than previously. Our main goal is to determine the informativeness of eye movements for task and mental state decoding. We perform two experiments. In the first experiment, we reanalyze the data from a previous study by Greene et al. (2012) and contrary to their conclusion, we report that it is possible to decode the observer's task from aggregate eye-movement features slightly but significantly above chance, using a Boosting classifier (34.12% correct vs. 25% chance level; binomial test, $p = 1.0722e - 04$). In the second experiment, we repeat and extend Yarbus's original experiment by collecting eye movements of 21 observers viewing 15 natural scenes (including Yarbus's scene) under Yarbus's seven questions. We show that task decoding is possible, also moderately but significantly above chance (24.21% vs. 14.29% chance-level; binomial test, $p = 2.4535e - 06$). We thus conclude that Yarbus's idea is supported by our data and continues to be an inspiration for future computational and experimental eye-movement research. From a broader perspective, we discuss techniques, features, limitations, societal and technological impacts, and future directions in task decoding from eye movements.

Introduction

Eyes are windows to perception and cognition. They convey a wealth of information regarding our mental processes. Indeed this has been elegantly demonstrated by seminal works of Guy T. Buswell (1935) and Yarbus (1967), who were the first to investigate the relationship between eye-movement patterns and high-level cognitive factors. Yarbus recorded observers' eye movements (with his homemade gaze tracking suction cap device) while they viewed the I. E. Repin painting, *The Unexpected Visitor* (1884).[1] He illustrated fixations of the observers as they viewed the painting under seven different instructions: (a) free examination, (b) estimate the material circumstances of the family, (c) give the ages of the people, (d) surmise (guess) what family had been doing before the arrival of the unexpected visitor, (e) remember the clothes worn by the people, (f) remember positions of people and objects in the room, and (g) estimate how long the visitor had been away from the family.

Yarbus's results show striking differences in eye-movement patterns across instructions over the same visual stimulus. Early in the viewing period, fixations were particularly directed to the faces of the individuals in the painting and observers showed a strong preference to look at the eyes more than any other features of the face. Yarbus concluded that the eyes fixate on those scene elements that carry useful information, thus showing where we look depends critically on our cognitive task. Further, Yarbus's experiments point towards the active nature of the human visual system as opposed to passively or randomly sampling the visual environment. This active aspect of vision and attention has been extensively

investigated by Dana Ballard, Mary Hayhoe, Michael Land, and others who studied eye movements in the context of natural behavior. Please see Ballard, Hayhoe, and Pelz (1995); Borji and Itti (2013); Hayhoe and Ballard (2005); Itti and Koch (2001); Land (2006); Land and Hayhoe (2001); Navalpakkam and Itti (2005); Schütz, Braun, and Gegenfurtner (2011); Tatler, Hayhoe, Land, and Ballard (2011); Tatler and Vincent (2009) for recent reviews.

Two prominent yet contrasting hypotheses attempt to explain eye movements and attention in natural behavior. First, according to the cognitive relevance hypothesis, eyes are driven by top-down factors that intentionally direct fixations toward informative task-driven locations (e.g., in driving). Second, in the absence of such task demands (e.g., in scene-free viewing), eyes are directed to low-level image discontinuities such as bright regions, edges, colors, etc., so-called salient regions. This is often referred to as the saliency hypothesis (Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; Treisman & Gelade, 1980). Both hypotheses are likely to be correct, yet the relative contribution of top-down and bottom-up attentional components varies across daily behaviors. Conversely, by looking at eye movements, one could possibly infer the underlying factors affecting fixations (i.e., task at hand or mental state) or gain insights into what an observer is currently thinking. Active research is undergoing to discover the interplay between top-down task-driven factors and bottom-up stimulus-driven factors in driving visual attention and to assess the amount of information eye movements convey regarding mental thoughts.

Yarbus showed a proof of concept with a single observer but did not conduct a comprehensive quantitative analysis. Perhaps DeAngelus and Pelz (2009) were the first to confirm Yarbus's findings, with multiple observers viewing Repin's painting. Viewing times in their study were self-paced (9–50 s), and were significantly less than the enforced 3-min viewing time of Yarbus's observer. DeAngelus and Pelz showed that observers' eye-movement patterns were similar to those reported by Yarbus, with faces invariably fixated and the overall viewing pattern varying with task instruction. A few of their observers, especially those with shorter viewing times, did not show dramatic shifts with instruction. The task "Give the ages of the people" resulted in the smallest interobserver distance of all tasks, indicating that for this task the eye-movement patterns were most similar among the observers. The "Estimate how long the visitor had been away from the family" task showed the most variability among observers, suggesting that observers used different viewing strategies to complete this task.

The general trend for fixations when viewing scenes to fall preferentially on persons within the scene had been shown previously by Buswell (1935). The tendency of observers to fixate on faces has recently been quantitatively confirmed by Cerf, Frady, and Koch (2009) and further supported by large-scale eye-tracking studies (e.g., Judd, Ehinger, Durand, & Torralba, 2009; Subramanian, Sebe, Kankanhalli, & Chua, 2010). Yarbus's results (along with DeAngelus & Pelz, 2009) indicate that, for extended viewing times, observers show a clear tendency to make repeated cycles of fixations between the key features of a face or a scene (i.e., cyclic behavior). Both attention and face perception communities have been largely inspired by Yarbus's early insights (see Kingstone, 2009).

Castelhano, Mack, and Henderson (2009) investigated how task instruction influences specific parameters of eye-movement control. They asked 20 participants to view color photographs of natural scenes under two instruction sets: searching a scene for a particular item or remembering characteristics of that same scene. They found that viewing task biases aggregate eye-movement measures such as average fixation duration and average saccade amplitude. Mills, Hollingworth, Van der Stigchel, Hoffman, and Dodd (2011) examined the influence of task set on the spatial and temporal characteristics of eye movements during scene perception. They found that task affects both spatial (e.g., saccade amplitude) and temporal characteristics of fixations (e.g., fixation duration).

Tatler, Wade, Kwan, Findlay, and Velichkovsky (2010) explored Yarbus's biography, his scientific legacy including his eye tracking apparatus, and his key contributions. They recorded eye movements of observers when viewing Yarbus's own portrait under the task conditions resembling Yarbus's questions with mild modifications. For example Questions 4 and 7 were phrased as "Estimate what the person had been doing just before this picture was taken" and "Try to estimate how long this person had been away from home when this picture was taken and why he had been away," respectively. They showed that: (a) Yarbus's findings generalize to a simpler visual stimulus and (b) instructions influence where and which features an observer inspects in face viewing.

Betz, Kietzmann, Wilming, and König (2010) addressed how or whether at all high-level task information interacts with the bottom-up processing of stimulus-related information. They recorded viewing behavior of 48 observers on web pages for three different tasks: free viewing, content awareness, and information search. They showed that task-dependent differences in their setting were not mediated by a reweighting of features in the bottom-up hierarchy, ruling out the weak top-down hypothesis. Consequently, they concluded that the strong top-down

hypothesis, which proposes that top-down information acts independently of the bottom-up process, is the most viable explanation for their data. These results support Yarbus's findings in that top-down factors influence where we look when viewing a scene.

Henderson, Shinkareva, Wang, Luke, and Olejarczyk (2013) recorded eye movements of 12 participants while they were engaged in four tasks over 196 scenes and 140 texts: scene search, scene memorization, reading, and pseudo reading. They showed that the viewing tasks were highly distinguishable based on eye-movement features in a four-way classification. They reported a high task decoding accuracy above 80% using multivariate pattern analysis (MVPA) methods widely used in the neuroimaging literature. Their four tasks, however, are much coarser than Yarbus's original questions, thus making the decoding problem effectively easier. Further, natural scenes and text used by Henderson et al. have dramatically different low-level feature distributions, which causes major differences in eye-movement patterns (Harel, Moran, Huth, Einhaeuser, & Koch, 2009; O'Connell & Walther, 2012), hence some of the decoding accuracy may be due to stimulus rather than task.

The list of studies addressing task decoding from eye movements and effects of tasks/instructions on fixations is not limited to the above works. Indeed, a large variety of studies has confirmed that eye movements contain rich signatures of the observer's mental task, including: predicting search target (Haji-Abolhassani & Clark, 2013; Rajashekar, Bovik, & Cormack, 2006; Zelinsky, Peng, & Samaras, 2013; Zelinsky, Zhang, & Samaras, 2008), predicting stimulus category (Borji, Tavakoli, Sihite, & Itti, 2013; Harel et al., 2009; O'Connell & Walther, 2012), predicting what number a person may randomly pick (Loetscher, Bockisch, Nicholls, & Brugger, 2010), predicting mental abstract tasks (Brandt & Stark, 1997; Ferguson & Breheny, 2011; Mast & Kosslyn, 2002; Meijering, van Rijn, Taatgen, & Verbrugge, 2012), predicting events (Bulling, Ward, Gellersen, & Tröster, 2011; Jang, Lee, Mallipeddi, Kwak, & Lee, 2011; Peters & Itti, 2007), classifying patients from controls (Jones & Klin, 2013; Tseng et al., 2012), and predicting driver's intent (Cyganek & Gruszczynski, 2014; Lethaus, Baumann, Köster, & Lemme, 2013). Several studies have investigated the role of eye movements in natural vision including: reading (Clark & O'Regan, 1998; Kaakinen & Hyönä, 2010; Rayner, 1979; Reichle, Rayner, & Pollatsek, 2003), visual search (Torralba, Oliva, Castelhano, & Henderson, 2006; Zelinsky, 2008), driving (Land & Lee, 1994; Land & Tatler, 2001), tea making (Land, Mennie, & Rusted, 1999), sandwich making (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003), arithmetic and geometric problem solving (Cagli et al., 2009; Epelboim & Suppes, 2001), mental imagery (Kosslyn, 1994; Mast & Kosslyn, 2002),

cricket (Land & McLeod, 2000), fencing (Hagemann, Schorer, Canal-Bruland, Lotz, & Strauss, 2010), billiard (Crespi, Robino, Silva, & deSperati, 2012), drawing (Coen-Cagli et al., 2009), magic (Kuhn, Tatler, Findlay, & Cole, 2008; Macknik et al., 2008), shape recognition (Renninger, Coughlan, Verghese, & Malik, 2004), and walking and obstacle avoidance (Mennie, Hayhoe, & Sullivan, 2007).

Departing from the above studies arguing that it is possible to decode observers' task from fixations (e.g., Henderson et al., 2013; Iqbal & Bailey, 2004), Greene, Liu, and Wolfe (2012) recently cast a shadow on task or mind state decoding by bringing counter examples. They conducted an experiment in which they recorded eye movements of observers when viewing scenes under four highly overlapped questions. Using three pattern classification techniques they were not able to decode the task significantly above chance using aggregate eye-movement features (see figure 4 in Greene et al.'s, 2012, paper). They were, however, able to decode image and observer's identity from eye movements above chance level. Task classification failure along with their finding that human judges could not tell the category of scan path, led Greene et al. to conclude "We have sadly failed to find support for the most straight-forward version of this compelling claim (Yarbus' claim). Over the range of observers, images and tasks, static eye movement patterns did not permit human observers or pattern classifiers to predict the task of an observer" (p. 7).

In summary, the effect of task on eye-movement patterns has been confirmed by several studies. Despite the volume of attempts at studying task influences on eye movements and attention, fewer attempts have been made to decode observer's task, especially on complex natural scenes using pattern classification techniques (i.e., the reverse process of task-based fixation prediction). However, there is of course a large body of work examining top-down attentional control and eye movements using simple stimuli and tasks such as visual search arrays and cueing tasks (e.g., Bundesen, Habekost, & Kyllingsbœk, 2005; Duncan & Humphreys, 1989; Egeth & Yantis, 1997; Folk & Remington, 1998; Folk, Remington, & Johnston, 1992; Sperling, 1960; Sperling & Dosher, 1986; Yantis, 2000). We attempt to thoroughly investigate the task decoding problem by analyzing previous data and findings of Greene et al. (2012) as well as our own collected data. We focus on Greene et al.'s study because we believe that their experimental design was best suited for task decoding and well in line with Yarbus's original idea, yet they reported that decoding failed. Further, we study limitations and important factors in task decoding including features and methods used for this purpose. Finally, we discuss potential technological and societal impacts of task and mental state decoding.

Images 1–5	6–10	11–15	16–20
4 O × T 1	4 O × T 2	4 O × T 3	4 O × T 4
4 O × T 2	4 O × T 3	4 O × T 4	...
5 O × T 3	5 O × T 4	...	
4 O × T 4	4 O × T 1		

Table 1. Arrangement of observers over tasks in Greene et al. (2012). O and T stand for observer and task, respectively.

Experiment 1

Due to important implications of Greene et al.'s (2012) results, here we first reanalyze their data and then summarize learned lessons. They shared data of their third experiment with us, which includes fixations of 17 observers viewing 20 grayscale images, each for 60 s.[2] They asked observers to view images under four questions: (a) Memorize the picture (memory), (b) determine the decade in which the picture was taken (decade), (c) determine how well the people in the picture know each other (people), and (d) determine the wealth of the people in the picture (wealth). Table 1 shows the arrangement of observers over these tasks. Each observer did all four tasks but over different images. This results in $17 \times 20 = 340$ scan paths where each scan path contains fixations of an observer over one image. The design was intentional to prohibit one observer seeing the same scene twice. Figure 1 demonstrates the stimuli used in this experiment.

Three factors may have caused task prediction failure in Greene et al.'s (2012) study: First and foremost, spatial image information is lost in the type of features they exploited (i.e., using histograms). This is particularly important since the first observation that strikes the mind from Yarbus's illustration is spatial patterns of fixations.[3] Second, the importance of the classification technique may have been underestimated in Greene et al.'s study. In fact, they only tried linear classifiers (linear discriminant analysis, linear support vector machine [SVM], and correlational methods) and concluded that their failure in task decoding is independent of the classification technique. They made similar arguments for images and features. Third, in Greene et al.'s study, observers were partitioned across images. Thus image and observer idiosyncrasies might have effects on task decoding (Chua, Boland, & Nisbett, 2005; Poynter, Barber, Inman, & Wiggins, 2013; Risko, Anderson, Lanthier, & Kingstone, 2012). For example, one observer might not have the necessary knowledge regarding a task or an image may not convey sufficient information for answering questions. In what follows, we scrutinize these factors one by one.

Regarding the first factor, we use a simple feature that is the smoothed fixation map, down sampled to 100×100 and linearized to a $1 \times 10,000$ D vector (Feature Type 1). Figure 2A shows fixation maps for an example image. The fixation map reflects pure eye-movement patterns. Additionally, we use histograms of normalized scan path saliency (NSS) proposed by proposed by Peters, Iyer, Itti, and Koch (2005), using nine saliency models.[4] This feature reflects the *stimulus + behavior* effect and basically indicates which visual attributes may be important when an observer is viewing an image under a task. NSS values are activations at fixated locations from a saliency map that is normalized to have zero mean and unit standard deviation. For each image, NSS values are calculated and then the histogram of these values (using 70 bins) is considered as features. Thus for nine models, this leads to a $9 \times 70 = 630$ dimensional vector (Feature Type 2). Although our features are aggregates and histograms like Greene et al.'s (2012), one critical difference is that the values that are aggregated reflect a spatial correlation between eye movements and spatial saliency features in each image. Thus, our features capture whether a task may lead an observer to allocate gaze differently over different types of salient image regions.

We also consider the first four features used in Greene et al. (2012) including the number of fixations, the mean fixation duration, the mean saccade amplitude, and the percent of the image area covered by fixations assuming a $1°$ fovea (Feature Type 3; dimensionality of four). In addition, because it has been argued that the first few fixations over a scene may convey more information (Parkhurst et al., 2002), we form a fourth feature type that includes $< x, y >$ locations of the first five fixations (i.e., a 10D vector). Note that, in addition to these features, one could think of more complex features (e.g., scan path sequence, NSS histograms on learned top-down task relevance maps, dwell times on faces, text, and human bodies, and temporal characteristics of fixations, Mills et al., 2011) to obtain better accuracies. But as we show here, these simple features suffice to decode the task in this particular problem.

Regarding the second factor, we investigate other classification methods such as k-nearest-neighbor; kNN (Fix & Hodges, 1951) and boosting (Freund & Schapire, 1997; Schapire, 1990) techniques that have been proven to be successful on different problems in machine learning, computer vision, and cognitive sciences.[5] The intuition is that for different problems, different classification methods may perform better. kNN is a classic nonparametric method for classification and regression problems. Given a distance metric (e.g., Euclidean distance), the kNN classifier predicts class label of a test sample as the majority vote of its k closest training examples in the feature space (i.e., the most common output among the neighbors). If $k = 1$, then the class label of the test sample is the same as its nearest neighbor. We also tried boosting algorithms

Easy   Hard

image 1 - 0.4906 [0.2500]   2 - 0.3741 [0.3088]   3 - 0.3518 [0.2537]   4 - 0.2812 [0.2279]

5 - 0.3094 [0.5331]   6 - 0.3565 [0.2610]   7 - 0.2694 [0.3492]   8 - 0.5118 [0.2941]

9 - 0.4835 [0.1728]   10 - 0.2800 [0.2206]   11 - 0.3977 [0.1765]   12 - 0.3965 [0.5845]

13 - 0.1859 [0.2280]   14 - 0.3623 [0.5956]   15 - 0.3765 [0.4118]   16 - 0.3600 [0.3382]

17 - 0.2259 [0.1765]   18 - 0.3270 [0.4154]   19 - 0.1318 [0.4118]   20 - 0.3670 [0.3235]

Figure 1. Stimuli used in Experiment 1. Easy and difficult scenes for task decoding are marked with blue and red boxes, respectively. Please see Appendix 1 for performances of individual runs of the RUSBoost classifier. Average decoding accuracies (numbers after dash lines) are using Feature Type 3 over 50 RUSBoost runs. Numbers in brackets are classification accuracy using Feature Type 1 (over 50 RUSBoost runs). Original images are 800 × 600 pixels.
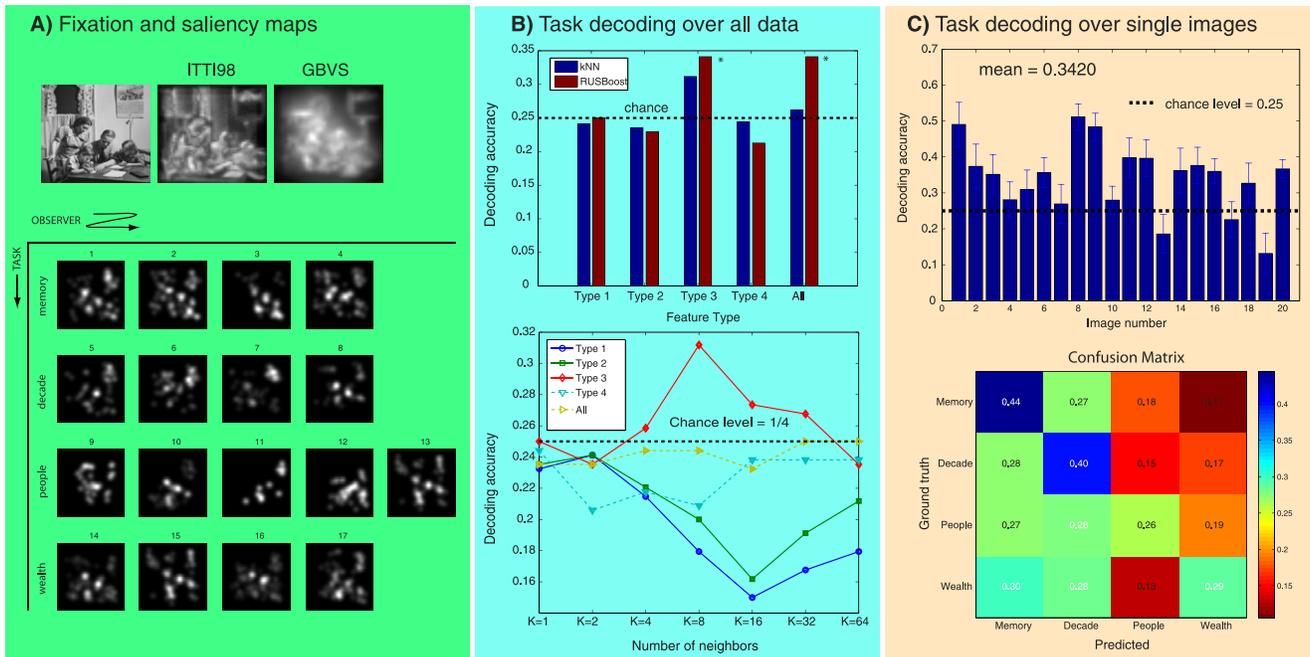
Figure 2. Results of Experiment 1: (A) Top: A sample image along with saliency maps using ITTI98 and GBVS models and its corresponding smoothed fixation maps (using Gaussian sigma 33 subtending about 0.85° × 0.85° of visual angle). Matlab code for generating the smoothed fixation map: imresize(conv2(map, fspecial('gaussian', 200, 33)), [100 100], 'nearest'). Numbers on top of fixation maps in the bottom panel show the observer's number (see Table 1). (B) Top: Task decoding accuracy using individual features and their combination over all data. Stars indicate statistical significance versus chance using binomial test. Bottom: Effect of number of kNN neighbors on task decoding accuracy. (C) Top: Average decoding accuracies over 50 runs of the RUSBoost classifier over individual images using Feature Type 3 (see Appendix 1). Error bars indicate standard deviations over 50 runs. Bottom: Average confusion matrix (over 50 RUSBoost runs) averaged over all images.

that are popular and powerful machine learning tools nowadays. The basic idea underlying boosting algorithms is learning several weak classifiers (i.e., a classifier that works slightly better than chance) and combing their outputs to form a strong classifier (i.e., a meta-algorithm). The learning is done in an iterative manner. After adding a weak learner, the data is reweighted to emphasize mistakes. Misclassified exemplars gain higher weight while correctly classified exemplars lose weight. Here, we employ the RUSBoost (random undersampling boost) algorithm (Seiffert, Khoshgoftaar, Van Hülse, & Napolitano, 2010), which uses a hybrid sampling/boosting strategy to handle

class imbalance problem in data with discrete class labels. To better model the minority class, this algorithm randomly removes examples from the majority class until all classes have balanced number of examples (i.e., undersampling). Due to the random sampling, different runs of this algorithm may yield different results. While the class imbalance (only one task has five subjects) is not a big issue in our data, we believe it is the ensemble of weak classifiers (here decision trees) that makes good prediction possible.[6]

With respect to the third factor, we conduct the following two analyses: (a) pooling data from all observers over all images and tasks (i.e., 17 × 20 scan paths) and (b) treating each image separately. These analyses help disentangle the effects of image and observer parameters on task decoding.

Task decoding over all data

We trained multiclass classifiers to recover task (one out of four possible) from eye-movement patterns. We follow a leave-one-out cross validation procedure similar to Greene et al. (2012). Each time we set one data point aside and train a classifier over the rest of

Images 1–5	6–10	11–15
3 O × T 1	3 O × T 2	3 O × T 3
3 O × T 2	3 O × T 3	3 O × T 4
3 O × T 3	3 O × T 4	3 O × T 5
3 O × T 4	3 O × T 5	...
3 O × T 5	...	...
3 O × T 6	...	...
3 O × T 7	3 O × T 1	3 O × T 2

Table 2. Arrangement of observers over tasks in Experiment 2. O and T stand for observer and task, respectively.

data. The trained classifier is then applied to the set-aside data point. We repeat the same procedure over all 340 data points and report the average accuracy (i.e., 340 binary values). Decoding results are shown in Figure 2B.

Using kNN and Feature Type 1 (i.e., fixation map), we achieved average accuracy of 0.2412 ($k = 2$; binomial test, $p = 0.89$). Feature Type 2 leads to accuracy of 0.2353 ($k = 2$; $p = 0.31$). Using Feature Type 3, we achieved accuracy of 0.3118, which is above Greene et al.'s (2012) results and is significantly better than chance ($k = 8$; $p = 0.014$). Classification with Feature Type 4 leads to accuracy of 0.2441 ($k = 1$). Combination of features did not improve the results significantly. Figure 2B (bottom panel) shows kNN performance as a function of number of neighbors ($k = 2^n$, $n = 0 \ldots 6$). kNN classification performance levels here are for the best-performing value of k.

Using RUSBoost classifier with 50 boosting iterations and Feature Type 1, we achieved accuracy of 0.25 (nonsignificant vs. chance; binomial test, $p = 0.6193$). We achieved accuracy of 0.2294 using Feature Type 2. Feature Type 3 leads to accuracy of 0.3412 ($p = 1.0722e - 04$). Finally, Feature Type 4 results in accuracy of 0.2176. Combination of all features did not increase the results significantly (accuracy of 0.3412 using all features).

Bonferroni correction for multiple comparisons (Shaffer, 1995): Here, we used two classifiers, five feature types (including combination of features), and seven values of parameter k in kNN resulting in $5 \times 7 + 5 = 40$ tries. We need to correct p values for these comparisons. Thus, significance p is equal to $0.05/40 = 0.0013$. Using kNN, the best p value is 0.014, which is above the corrected significance level; therefore kNN does not yield statistically significant decoding accuracy that is strong enough given that we made 40 different attempts at decoding. Hence, we discard using kNN for the rest of the paper. Using the RUSBoost classifier (with Feature Type 3), however, results remain statistically significant after correction as p values are smaller than 0.0013, which indicates that task is decodable on this data significantly above the 25% chance level.

Results of this analysis indicate that spatial fixation patterns are not informative regarding the observer's task when pooling all data (on Greene et al.'s, 2012, data). Further, our results show that classification method is a key factor. For example, using the same four features employed by Greene et al. (2012) (Feature Type 3), we achieved better accuracies with kNN and boosting classifiers. Note that here we did not conduct an exhaustive search to find the best features or feature combinations. It might be possible to reach even higher accuracies with more elaborate feature selection strategies.

Task decoding over single images

Task decoding accuracy highly depends on the stimulus set. For example, if an image does not have the necessary content that is called for by different tasks (in an extreme case, a blank image and tasks about age or wealth of people), it may not yield task-dependent eye-movement patterns as strong as an image that has such content. That is, we expect that interaction between semantic image contents and task may give rise to the strongest eye-movement signatures. Failure to decode task might thus be more likely if the stimuli do not support executing the task. This is particularly important since both Yarbus and Greene et al. (2012) did not probe observers' responses to see whether or not they were actually able to perform the task.

We train a RUSBoost classifier (with 50 boosting iterations) on 16 observers over each individual image and apply the trained classifier to the remaining observer over the same image (i.e., leave one observer out). We repeat this process for all 20 images. Using Feature Type 1, we achieve average accuracy of 0.3267 (over 50 runs and images). Feature Type 3 resulted in accuracy of 0.3414 (see Appendix 1, for results of 50 runs). The maximum performance using this feature over runs was 0.3719 and the minimum was 0.3156. Using combination of all features (a feature vector of size $10,000 + 9 \times 70 + 4 + 10 = 10,644$ D) results in average accuracy of 0.3294. Examination of confusion matrices using RUSBoost and Feature Type 3 (Figure 2C) shows above chance performance on diagonal elements with higher accuracies for memory and decade tasks. There is high confusion between wealth and other classes.

Average task decoding performance per image using Feature Type 3 is illustrated in Figure 2C as well as in Figure 1. Using this feature, decoding accuracy is significantly above chance level for majority of the images, is nonsignificant versus chance for one image, and is significantly below chance for three images (using t test over 50 runs; see Appendix 1). The easiest and most difficult stimuli using Feature Type 1 along with their scan paths and confusion matrices (using a sample run of RUSBoost) are shown in Figure 3.

Results of the second analysis support our argument that image content is an important factor in task decoding. Task decoding becomes very difficult if an image lacks diagnostic information relevant to the task (see Figure 3 for such an example). Further, when treating each image separately, scan paths turn to become informative regarding the task for some images. Overall, results from this experiment suggest that it is possible to decode the task above chance from the same type of features used by Greene et al. (2012).
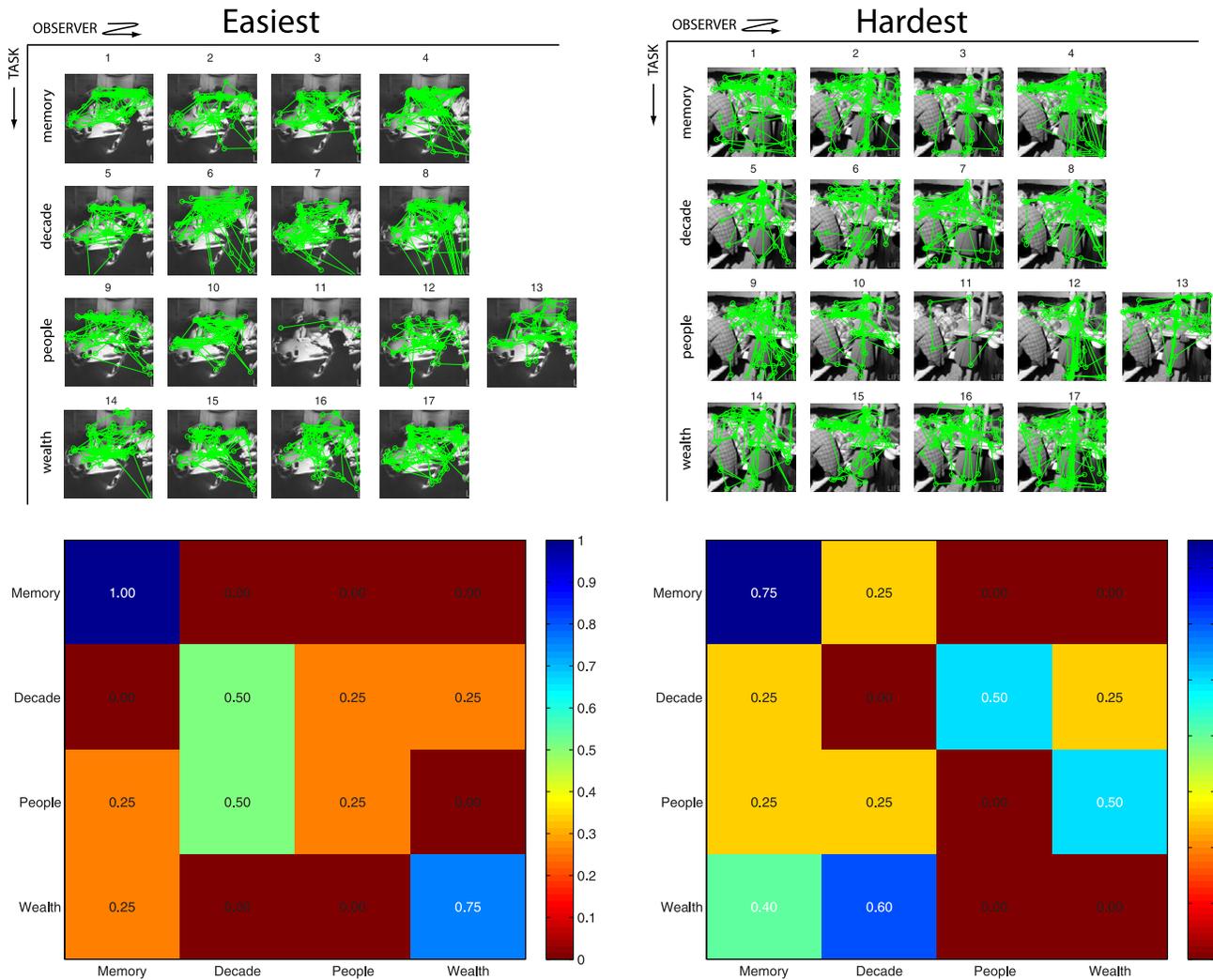
Figure 3. Easiest and hardest stimuli for task decoding in Experiment 1 using Feature Type 1 over 50 RUSBoost runs. Confusion matrices are for a sample run of RUSBoost on each image using leave-one-out procedure.

Experiment 2

The questions in the task set of Greene et al. (2012) are very similar to each other. For decade, memory, and wealth tasks, observers have to look over the entire image to gain useful information. This causes very similar fixation patterns such that these patterns are not really differentiable to the naked eye (see Figure 3) or to classifiers, while, to some extent, clearly different patterns may have contributed to making Yarbus's illustration so compelling. Here we aim to decode observer's task from eye movements with particular emphasis on spatial fixation patterns (i.e., Feature Type 1) rather than aggregate features (Type 3). While mean values of eye-movement measures (i.e., Feature Type 3) can change as a function of task, the distributions of these values highly overlap across tasks (Henderson et al., 2013).

In our view an important limitation of Greene et al.'s study is that they did not use Yarbus's original seven tasks, as Yarbus might have reached different conclusions had he used different tasks. In this experiment, we thus seek to test the accuracy of Yarbus's exact idea by replicating his tasks.

Methods

Participants

A total of 21 students (10 male, 11 female) from the University of Southern California (USC) participated. Students' majors were computer sciences, neuroscience, psychology, mathematics, cognitive sciences, communication, health, biology, sociology, business, and public relations. The experimental methods were approved by the USC's Institutional Review Board (IRB). Observers had normal or corrected-to-normal vision and were compensated by course credits.
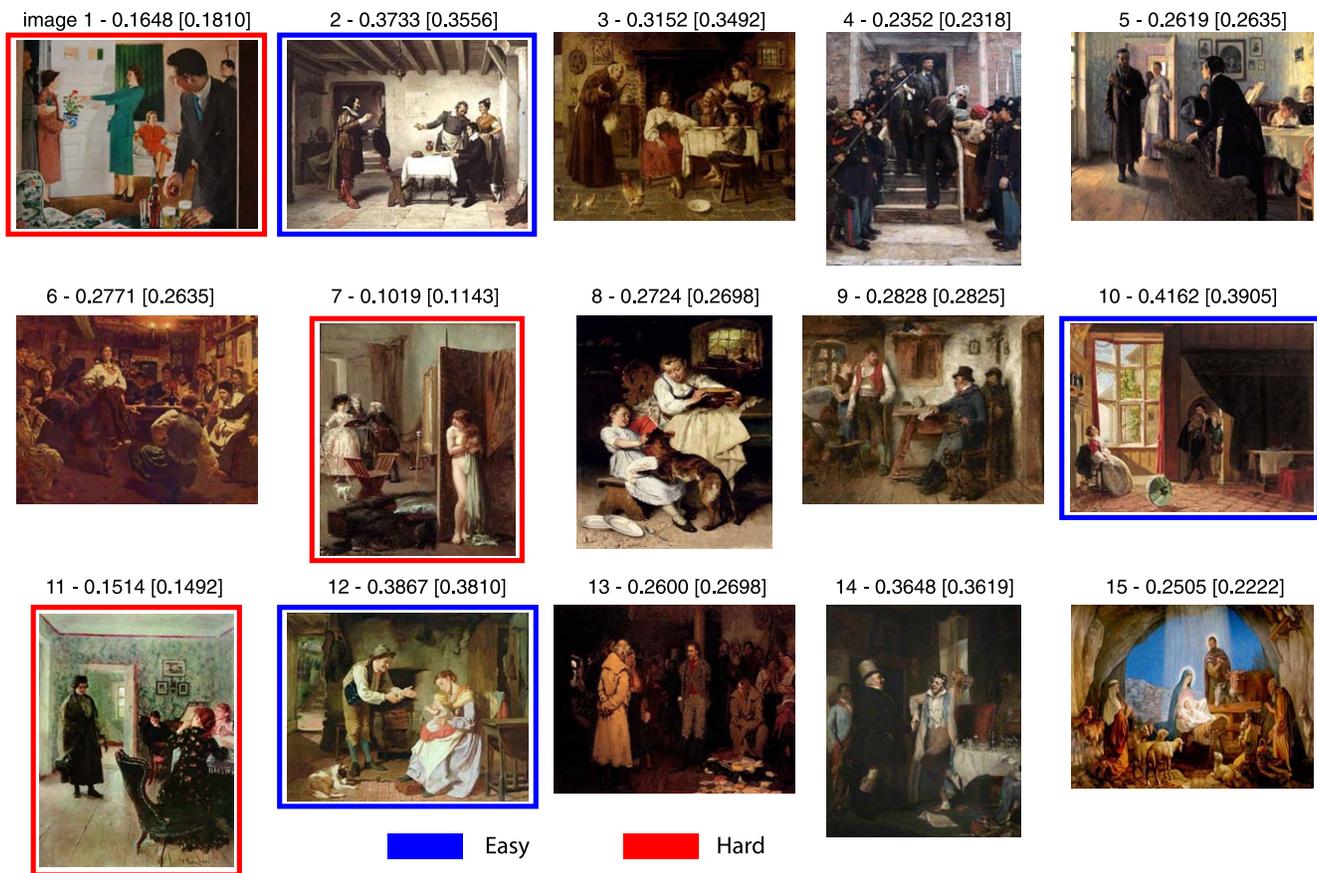
Figure 4. Stimuli used in Experiment 2. Images resemble Repin's painting (Image 5) in that in all of the images exists a somewhat unexpected visitor. (source: courtesy of http://www.ilyarepin.org). Three easiest and three most difficult stimuli are marked with blue and red boxes, respectively. Average decoding accuracies (numbers after dash lines) are using combination of Feature Types 1 and 2 over all RUSBoost runs. See Appendix 2 for decoding results on individual RUSBoost runs. Numbers in brackets are classification accuracy using Feature Type 1.

Observers were in the age range of 19–24 (mean = 22.2, SD = 2.6). They were naïve with respect to the purpose of the experiment.

Apparatus

Participants sat 130 cm away from a 42-in. monitor screen so that scenes subtended approximately 43° × 25° of visual angle. A chin/head rest was used to minimize head movements. Stimuli were presented at 60 Hz at resolution of 1920 × 1080 pixels. Eye movements were recorded via an SR Research Eyelink eye tracker (spatial resolution 0.5°) sampling at 1000 Hz.

Materials

Stimuli consisted of 15 paintings (13 are oil on canvas, some are by I. E. Repin). Figure 4 shows stimuli including Repin's painting used by Yarbus. We chose images such that a person[7] who could be construed as an unexpected visitor exists in all of them. Thus Yarbus's questions are applicable to these images (e.g., more so on Images 2, 3, and 11 and less so on Images 4, 6, and 15).

Procedure

We followed a partitioned experimental procedure similar to Greene et al. (2012), where observers answered questions on three sets of images (Table 2). Each set consists of five images corresponding to one row of Figure 4. In other words, no participant saw the same stimulus twice. Each image was shown for 30 s followed by a 5-s delay (gray screen). At the beginning of each session (five images), the eye tracker was recalibrated. Each observer viewed each set of five images only under one question. We used the seven questions of Yarbus's study mentioned in the Introduction. Figure 5 illustrates eye movements of observers on seven images.
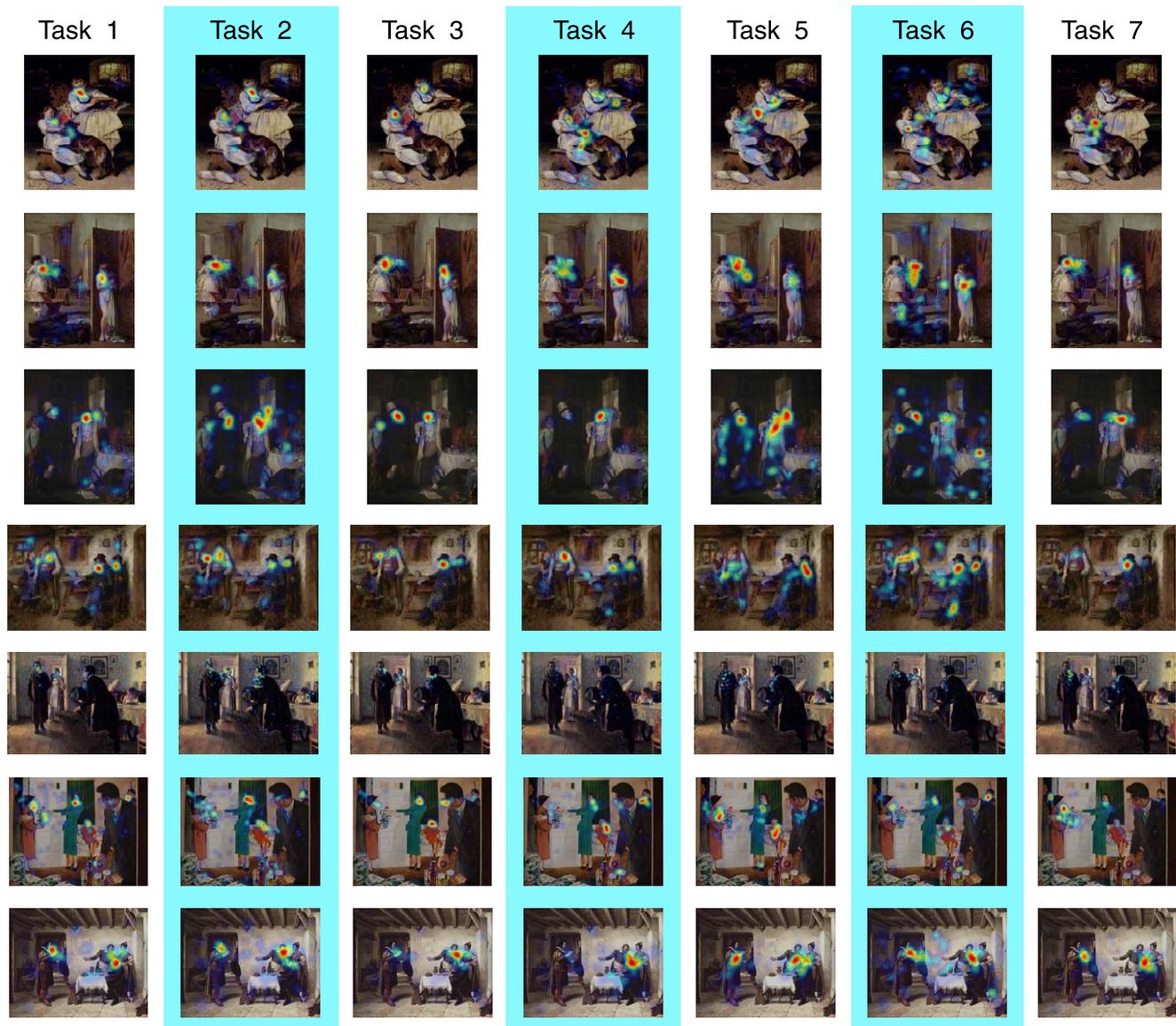
Figure 5. Eye movements of observers over stimuli in Experiment 2 for seven images. Note that each image was shown to a observer only under one question. Tasks are: (1) free examination, (2) give material circumstances (wealth), (3) estimate ages of the people, (4) estimate the activity before the arrival of the visitor, (5) remember clothes, (6) remember positions of people and objects, and (7) estimate how long the visitor had been away.

Decoding results

We employ the RUSBoost classifier with 50 boosting iterations as in the first experiment. Features consist of saliency maps of nine models used in Experiment 1 (Footnote 4) plus additional 14 feature channels from the ITTI model including: ITTI-C, ITTI-CIO, ITTI-CIOLTXE, ITTI-E, ITTI-Entropy, ITTI-I, ITTI, ITTI-L, ITTI-O, ITTI-OLTXE, ITTI-Scorr, ITTI-T, ITTI-Variance, and ITTI-X. These feature channels extract different types of features that range from intensity (I), color (C), orientation (O), entropy (E), variance, t-junctions (T), x-junctions (X), l-junctions (L), and spatial correlation (Scorr). Please see Itti et al.

(1998) and Tseng et al. (2012) (and its supplement) for more details on these features and implementation details. ITTI and ITTI98 are different versions of the Itti et al. model, corresponding to different normalization schemes. In ITTI98, each feature map's contribution to the saliency map is weighted by the squared difference between the globally most active location and the average activity of all other local maxima in the feature map (Itti et al., 1998). This gives rise to smooth saliency maps, which tend to correlate better with noisy human eye-movement data. In the ITTI model (Itti & Koch, 2000), the spatial competition for saliency is much stronger, which gives rise to much sparser
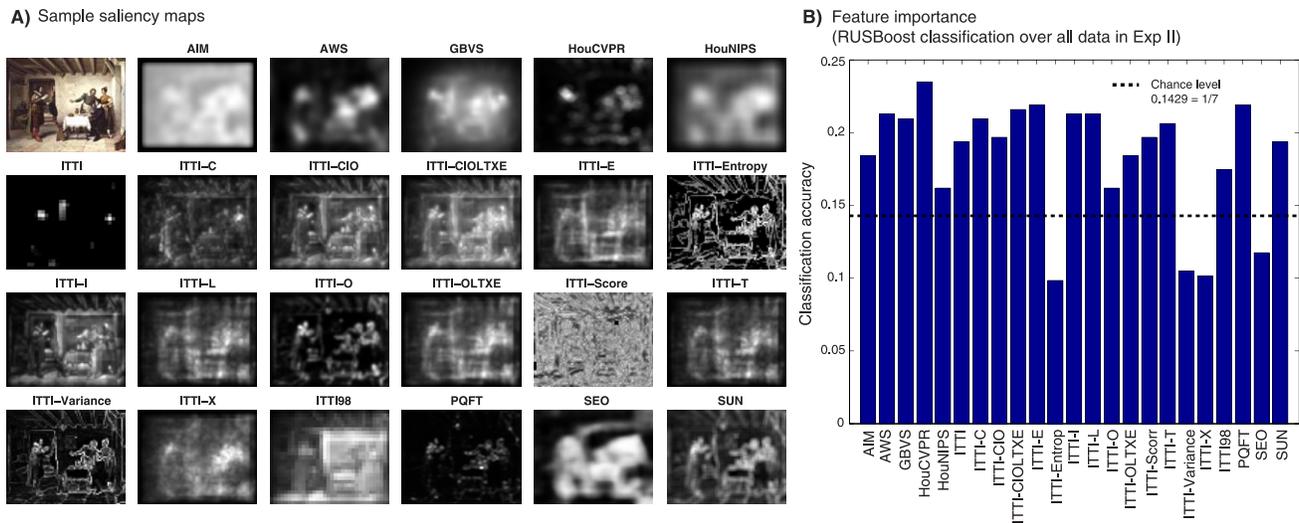
Figure 6. (A) Saliency maps for a sample image used in the second experiment. Acronyms are: intensity (I), color (C), orientation (O), entropy (E), variance, t-junctions (T), x-junctions (X), l-junctions (L), and spatial correlation (Scorr). (B) Importance of saliency maps (Feature Type 2 using 70D NSS histograms) for task decoding. Here, a RUSBoost classifier (50 runs) was used over all data according to the analysis in the section Task decoding over all data).

saliency maps. Figure 6A shows 23 saliency maps for a sample image.

Task decoding over all data: Following Experiment 1, we first pool all data and perform task decoding over all images and observers. We report results using a leave-one-out procedure. We have 21 observers, each viewing 15 images (each five images under a different question; three questions per observer) thus resulting in 315 scan paths. Using Feature Type 1, we achieved average accuracy of 0.2421, which is significantly above chance[8] (binomial test, $p = 2.4535e − 06$). Using Feature Type 2 (i.e., NSS histogram of nine saliency models as in Experiment 1) results in accuracy of 0.2254 ($p = 5.6044e − 05$). Increasing the number of saliency models to 23 results in the same performance as when using nine models. Combination of all features did not improve the results in this analysis. To evaluate the importance of saliency maps as features, we performed task decoding over all data using individual saliency features (i.e., NSS values; Feature Type 2) and RUSBoost classification (see the section Task decoding over all data). Results are shown in Figure 6B. Majority of the saliency models lead to above chance accuracy indicating informativeness of NSS histograms and low-level image features for task decoding.

Bonferroni correction for multiple comparisons: With the RUSBoost classifier, correcting for three features and their combination, p values have to be smaller than $0.05/4 = 0.0125$, which is the case here using all feature types. Thus, we can safely conclude that task is decodable from eye movements on our data using spatial fixation patterns and NSS histograms (as opposed to Experiment 1).

Task decoding over single images: Three observers viewed each image under one question thus resulting in 21 data points per image (i.e., 3 Observers × 7 Questions). Note that each set of three observers were assigned the same question (Table 2). RUSBoost classifier and Feature Type 1 results in average accuracy of 0.2724 over 50 runs and 15 images. Using first two feature types (a $10,000 + 23 × 70 = 11610D$ vector) results in average performance of 0.2743. Over all runs (i.e., table rows), the minimum accuracy (average over all 15 images) is 0.2540 and maximum accuracy is 0.3079. Note that our accuracies are almost two times higher than the 14.29% chance level (i.e., 1/7). Easy and difficult stimuli for task decoding are shown in Figure 4. See Appendix 2 for results of individual runs of the RUSBoost classifier over individual images.

To measure the degree to which tasks differ from each other, we show in Figure 7A the distribution of fixations over all images with the same task. Each element shows the amount of overlap in two questions. To generate this plot, we first normalize each map to [0 1] and then subtract maps from each other. Hence brighter blue and red regions mean higher difference between two tasks. It shows profound differences among Tasks 3 (estimating ages), 4 (estimating activity), and 7 (estimating away time) to other tasks. Task 1 (free examination) is more similar to other tasks. The reason might be because people look everywhere in images including faces and people, which are also informative objects for other tasks. Task 2 (estimating wealth) and Task 6 (remembering positions) show smaller difference to other tasks probably because observers inspect the entire image in two tasks. Figure 7B shows the confusion matrix averaged over 15

A) Similarity/difference of fixation maps   **B)** Confusion matrix of RUSBoost classifier in Exp II
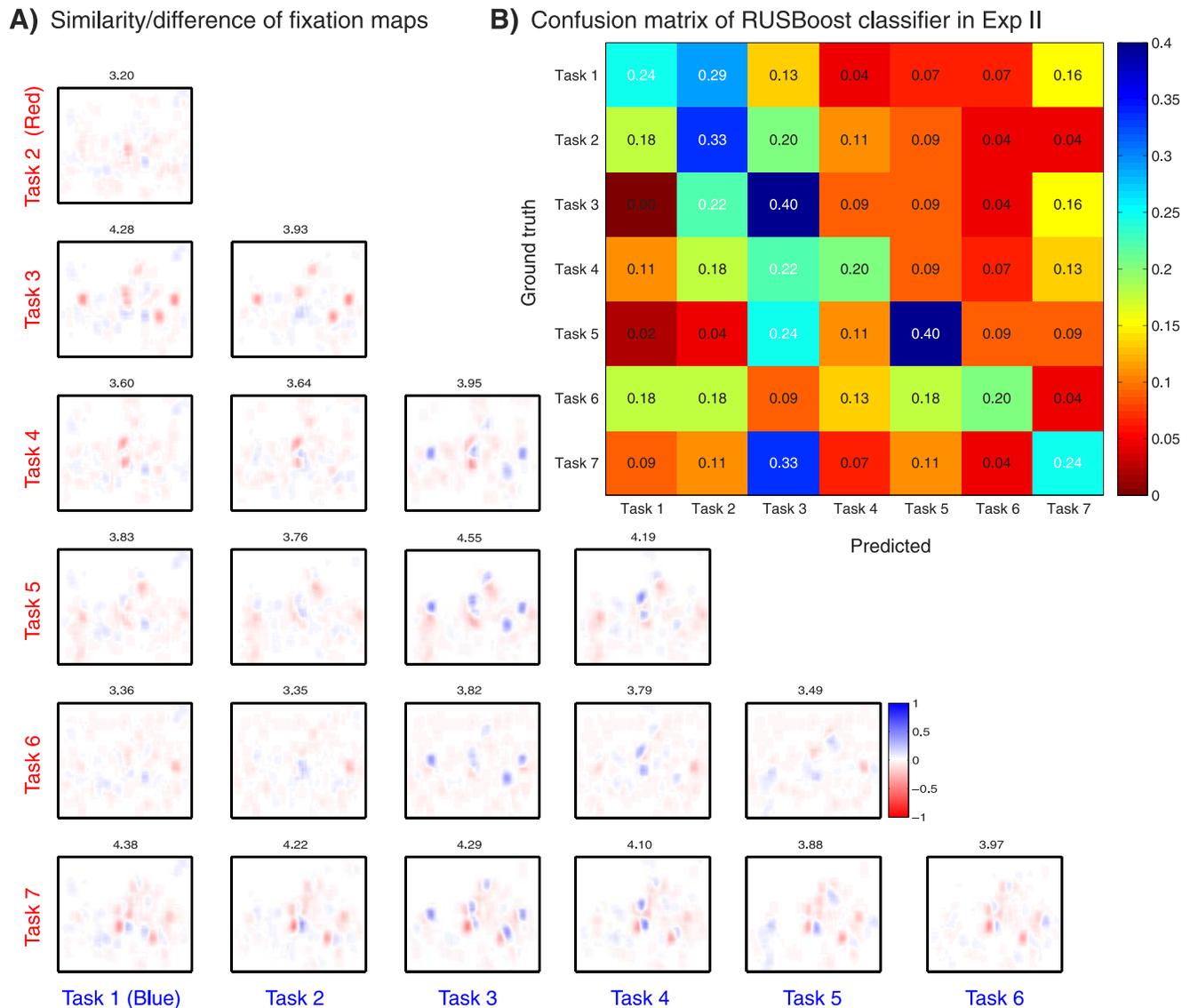
Figure 7. (A) Similarity/difference of tasks from human fixation maps in Experiment 2. Brighter red or blue regions mean higher difference. Values close to zero mean less difference. The numbers on top of each image show the sum of the absolute differences between two fixation maps. (B) Confusion matrix of the RUSBoost classifier averaged over 50 RUSBoost runs each on a single image using Feature Type 1.

images and 50 RUSBoost runs using Feature Type 1. We observe high accuracies for Task 3 (estimating age), Task 5 (remembering cloths), and Task 7 (estimating how long the visitor has been away) but low accuracy for the free-viewing task. There is a high confusion between Task 2 and Tasks 6 and 1 and also between Task 1 and Task 7. The easiest and hardest stimuli using Feature Type 1 along with their scan paths and confusion matrices are shown in Figure 8.

Results of the two analyses in second experiment, in alignment with DeAngelus and Pelz (2009), confirm that eye movements are modulated top down by task demands in a way that task can be predicted from eye-movement patterns. We found that spatial fixation patterns, which were not much informative over

Greene et al.'s (2012) data, suffice to decode the task on our data. We expect to gain even higher task decoding accuracies by using other eye-movement statistics, such as fixation durations or amplitudes, that have been shown to be different across Yarbus's questions (DeAngelus & Pelz, 2009).

Discussion and conclusion

What do we learn from the two experiments in this study? Successful task decoding results provide further evidence that fixations convey diagnostic information regarding the observer's mental state and task,[9]
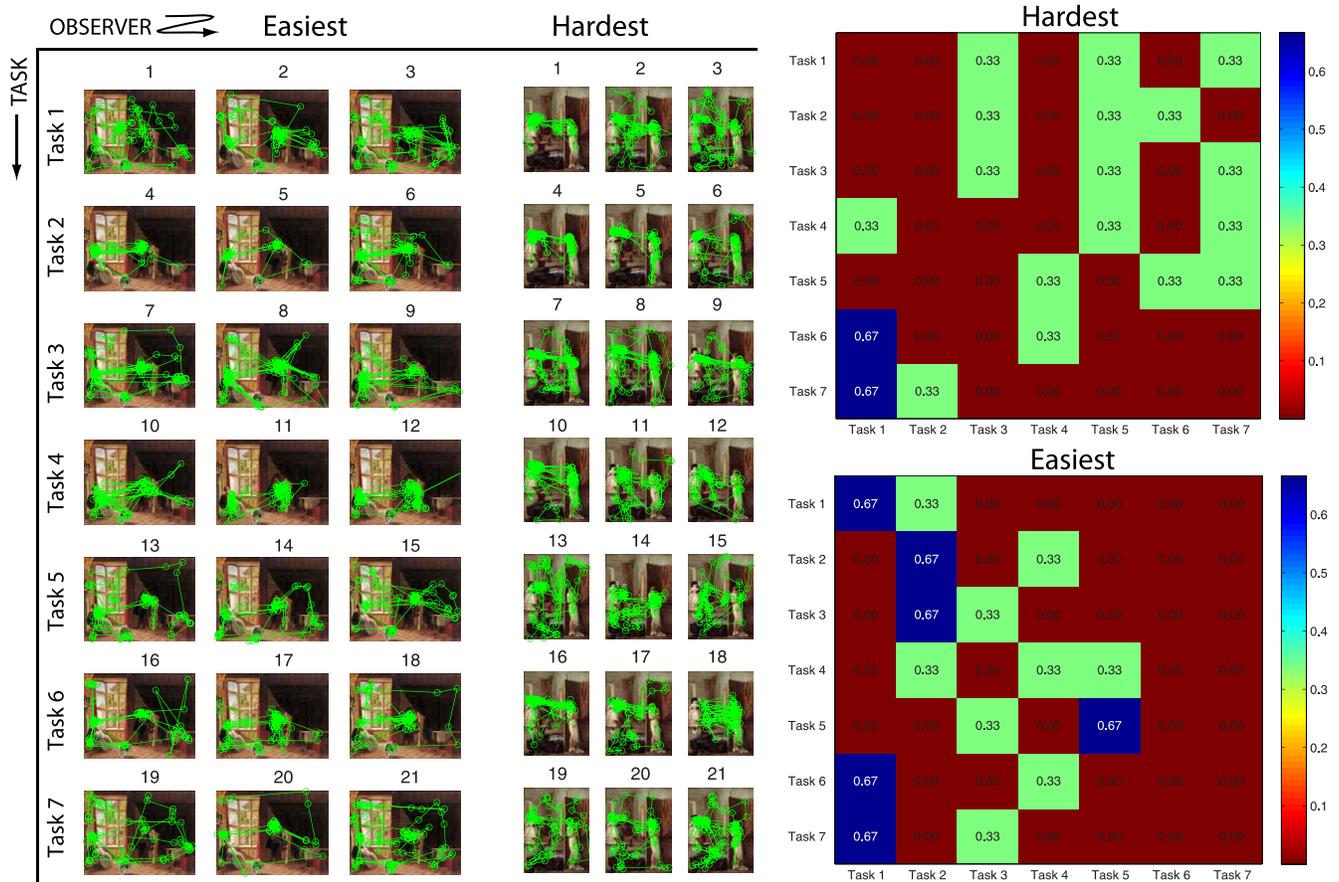
Figure 8. Easiest and hardest stimuli for task decoding in Experiment 2 using Feature Type 1 over 50 RUSBoost runs. Confusion matrices are for a sample run of RUSBoost on each image using leave-one-out procedure.

consistent with the cognitive relevance theory of attention (see Hayhoe & Ballard, 2005). This means that top-down factors in complex tasks systematically influence the viewer's cognitive state and his thought processes. Our results support previous decoding findings mentioned in the introduction section (e.g., some over more controlled stimuli such as predicting search target, Rajashekar et al., 2006).

We demonstrated that it is possible to reliably infer the observer's task from Greene et al.'s (2012) data using stronger classifiers. Classification was better when we treated images individually. Although we were able to decode the task from Greene et al.'s data, making strong arguments regarding feasibility of task decoding on this data is difficult mainly due to the small size of this dataset. We think to gain better insights, larger datasets for task decoding are necessary. Such datasets allow break down of data into (larger) separate train and test sets. Parameters of a classifier can be optimized using the train set and the resultant classifier can be evaluated on the test set. Performing the analysis in this manner eliminates the need for correction for multiple comparisons, hence allowing one to try possibly thousands of possible classifiers and parameters.

In the second experiment, we showed that it is possible to decode the task using Yarbus's original tasks, almost twice above chance, much better than using Greene et al.'s (2012) tasks. These results are in line with findings of DeAngelus and Pelz (2009). While our results are significantly above chance, it might be still possible to obtain better accuracies by exploiting even more informative features and other types of classification techniques. Our investigation on task decoding using 5-s time slots (i.e., first 5 s, second 5 s, . . .) suggest that accuracies might be higher for early fixations but this needs further investigation. We also found that decoding accuracy critically depends on three factors: (a) task set (how separable they are), (b) stimulus set (whether a scene has sufficient information or not), and (c) observer's knowledge (whether observers understand questions).

Just recently, we noticed that another group (Kanan et al., 2014) has been working on this problem in parallel. Using support vector machines (SVM) with radial-basis kernel function and C-SVC training algorithm and summary statistics features (a 2D vector comprised of mean fixation duration and the number of fixations in each trial), Kanan et al. achieved accuracy of 26.3% (95% CI = 21.4–31.1%, $p = 0.61$) which is not

significantly better than chance. However, using a SVM classifier with a Fisher Kernel Learning (FKL) algorithm with only motor information (i.e., fixation duration and location of each fixation; thus a variable number of 3D vectors in each trial) they were able to exceed chance level (33.1% correct, 95% CI = 27.9–38.3%). This analysis suggests that summary statistics alone are not enough for task decoding and it is necessary to add spatial information with the tasks and images of Greene et al. Further, they conducted a within-subjects analysis (i.e., training a classifier on each subject individually) following a leave one out procedure (thus repeating this procedure 20 times per subject). SVM (with summary statistics features) and SVM-FKL classifiers resulted in 38.8% accuracy (95% CI = 33.4–44.1%; chance = 25%) and 52.9% accuracy (95% CI = 46.4–57.4%), respectively. Overall, Kanan et al.'s result further support our findings here regarding availability of sufficient information diagnostic of task in Greene et al.'s data.

Is it always possible to decode task from eye movements? We argue that there is no general answer to this type of pattern recognition questions. Answers depend on the used stimuli, observers, and questions. One could choose tasks such that decoding becomes very hard even with sophisticated features and classifiers; we found that this is the case on Greene et al.'s (2012) data. In particular, on the type of tasks and scenes used here majority of fixations are attracted to faces and people, which causes a huge overlap across tasks. In some easier scenarios, where tasks are more different, very simple features might suffice to decode the task accurately (e.g., Henderson et al., 2013). In the extreme simplest case, one can imagine a task like this: a person on the left side of the screen and a dog on the right side with observers' tasks being: (a) How old is the person? and (b) what breed is the dog? Obviously answering these tasks demands looking at the person for the first question and looking at the dog for the second question, which results in 100% task decoding accuracy (for a rational observer) just from eye-movement locations. One can also choose images from which task decoding is very difficult because they contain little information that is directly relevant to the task. This was also found in our results, as some images yield more accurate task decoding than others. One could also recruit observers who don't understand the question. So far none of the works mentioned in the present study have analyzed the observers' answers on tasks. So, the failure in task decoding might be simply due to the observer's disability to extract useful information from the scene.

Since the parameter space is large, making strong arguments regarding impossibility of task decoding (see, e.g., Greene et al.'s, 2012, claim "static scan paths alone do not appear to be adequate to infer complex mental states of an observer" in their abstract) seems to be very difficult and needs a systematic probing of the whole parameter space (or a theoretical proof). On the other hand, to prove that task decoding on a particular setting is feasible, one only needs to find a working set of parameters (and it suffices; after accounting for multiple comparisons and following a cross-validation procedure, Salzberg, 1997). The latter is the common practice in pattern recognition community. Please note that our results also do not imply that it is always possible to decode the task. The counter example proposed by Greene et al. (2012) was found to not hold in our analysis.

As a control analysis, Greene et al. (2012) asked some human participants to look at eye movements of their observers and guess which task the observers have been doing. They showed that similar to classifiers, participants also failed in task decoding. Failure of humans to decode the task by looking at eye-movement patterns (experiment 4 in Greene et al., 2012) does not necessarily mean that fixations lack task-relevant information. Indeed, there are some cases in vision sciences where machine learning techniques outperform humans, in particular over large datasets (e.g., frontal face recognition, defect detection, cell type differentiation, DNA microarray analysis, etc.).

Several concerns exist that need to be carefully thought about before conducting a task decoding experiment using eye movements. Here we followed the procedure by Greene et al. (2012) in which: (a) no observer viewed the same image twice and (b) the same scene was shown under multiple questions. The first rule aims to eliminate memory biases. The second rule ensures that the final result is not due to differences in stimuli. DeAngelus and Pelz (2009) and Yarbus (1967) violated the first rule where the same observers viewed the images under the same questions. Henderson et al. (2013) violated the second rule in which different questions were asked over different images (which might be the reason why they obtained such high accuracies above 80%). Another possibly important factor affecting task decoding results is eye-tracking accuracy. This is particularly important when tasks are very similar to each other. One other concern regards selection of the stimulus set. If the stimulus set includes many images containing people, faces, and text, which capture a large portion of fixations in a task-independent manner, then there is basically not much information left helping task decoding. The last concern is about the suitability of features. In some scenarios, especially in dynamic environments (e.g., watching a video, driving a car, etc.) the type of features employed here may not be suitable for task decoding. In particular, spatial information is reduced to one fixation per frame. This requires temporal processing of features to see which places (or in what order) observers have visited the locations.

Here, we showed that task is decodable on static images by a more systematic and exhaustive exploration

of the parameter space including features, classifiers, and new data. Pushing deeper into real-time scenarios, using joint online analysis of video and eye movements, we have recently been able to predict—more than one second in advance—when a player is about to pull the trigger in a flight combat game, or to shift gears in a car racing game (Peters & Itti, 2007). We have been also able to predict next fixation of a video game player for such games as running a hot-dog stand (Borji, Sihite, & Itti, 2014) and Super Mario Cart (Borji, Sihite, & Itti, 2012a). In a similar approach where our computational models provide a normative gold standard against one particular individual's gaze behavior, we have demonstrated a system that can predict, by recording an observer's gaze for 15 min while one watches TV, whether one has ADHD (Tseng et al., 2012). These preliminary results clearly demonstrate how computational attention models can be used jointly with behavioral recordings to infer some internal state of a person, from a short-term intention (e.g., pull the trigger) to long-term characteristics (e.g., like ADHD).

Beyond scientific value, decoding task from eye movements has practical applications. Potential technological applications include: wearable visual technologies (smart glasses like Google Glass), smart displays, adaptive web search, marketing, activity recognition (Albert, Toledo, Shapiro, & Kording, 2012; Fathi, Farhadi, & Rehg, 2011; Pirsiavash & Ramanan, 2012), human–computer interaction, and biometrics. Portable electronic devices such as smartphones, tablets, and smart glasses with cameras are becoming increasingly popular (see Windau & Itti, 2013, for an example study). Enabling eye tracking on these devices could be used to predict the user's intent one step ahead and provide him necessary information in a more efficient and adaptive manner. This could be augmented with approaches that use nonvisual information on cell phones such as accelerometer data or global positioning systems (e.g., Albert et al., 2012). Another area of applicability is assistant systems especially for elderly and disabled users (e.g., in driving or other daily life activities (Bulling et al., 2011; Doshi & Trivedi, 2009, 2012). Here, we focused on predicting observer's task. Some studies have utilized eye movements to tap into mental states such as confusion and concentration (Griffiths, Marshall, & Richens, 1984; Victor, Harbluk, & Engström, 2005), arousal (Subramanian et al., 2010; Woods, Beecher, & Ris, 1978), or deception (Kuhn & Tatler, 2005). Eye movements can also be utilized as a measure of learning capacity in category learning and feature learning (e.g., Chen, Meier, Blair, Watson, & Wood, 2013; Rehder & Hoffman, 2005) and expertise (e.g., Bertram, Helle, Kaakinen, & Svedström, 2013; Jarodzka, Scheiter, Gerjets, & Van Gog, 2010; Vogt & Magnussen, 2007).

From a societal point of view, reliable fixation-based task decoding methods could be very rewarding. One area of application is patient diagnosis. Several high-prevalence neurological disorders involve dysfunctions of oculomotor control and attention, including Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Fetal Alcohol Spectrum Disorder (FASD), Parkinson's disease (PD), and Alzheimer's.[10] Diagnosis and treatment of these disorders are becoming a pressing issue in today's society (see Jones & Klin, 2013; Klin, Lin, Gorrindo, Ramsay, & Jones, 2009). For example, about one in six children in the United States had a developmental disability, such as intellectual disabilities, cerebral palsy, and autism, in 2006–2008. Reliable and early diagnosis of these disorders boils down to accessing observers' internal thought processes and their cognitive states. This is where our task decoding framework becomes relevant and could potentially replace or complement existing clinical neurological evaluation, structured behavioral tasks, and neuroimaging techniques that are currently expensive and time consuming. We believe that the type of methods discussed here, along with low-cost, noninvasive, eye-tracking facilities, offer considerable promise for patient screening. However, to make it happen in the future, high-throughput and robust task-decoding methods need to be devised. One direction could be augmenting eye movements with purely biological cues such as pupils, sweating, heart rate, and breathing for this purpose.

Keywords: visual attention, eye movements, bottom-up saliency, top-down attention, free viewing, visual search, mind reading, task decoding, Yarbus

Acknowledgments

Commercial relationships: none.
Corresponding author: Ali Borji.
Email: borji@usc.edu.
Address: Department of Computer Science, University of Southern California, Los Angeles, CA, USA.

Footnotes

[1]See Figure 4, Image 5.

[2]Please see the original paper for more details on the experimental setup. Greene et al. (2012) reported 16 observers on their paper (experiment 3) but shared 17 with us and on their website http://stanford.edu/~mrgreene/Publications.html. Our results and conclusions are valid over selection of 16 subjects distributed equally across tasks.

[3]Greene et al. (2012) were able to decode the stimulus from the aggregate features. We suspect that using spatial patterns will lead to much higher accuracies as scan paths on images are often quite different (e.g., Harel et al., 2009; O'Connell & Walther, 2012).

[4]Selected saliency models include: attention for information maximization (AIM) (Bruce & Tsotsos, 2009), adaptive whitening saliency (AWS) (Garcia-Diaz, Fdez-Vidal, Pardo, & Dosil, 2012), graph based visual saliency (GBVS) (Harel, Koch, & Perona, 2006), HouCVPR (Hou & Zhang, 2007), HouNIPS (Hou & Zhang, 2008), ITTI98 (Itti et al., 1998), phase spectrum of Quaternion Fourier transform (PQFT) (Guo & Zhang, 2010), SEO (Seo & Milanfar, 2009), and saliency using natural statistics (SUN) (Zhang, Tong, Marks, Shan, & Cottrell, 2008). For more details on these models, the interested reader is referred to Borji and Itti (2012) and Borji, Sihite, & Itti (2012b). Note that saliency is not a unique measurement and may change from one model to another. That is why here we employ several models instead of one.

[5]Boosting classifiers have been used for fixation prediction in free viewing tasks (e.g., Borji, 2012; Zhao & Koch, 2012).

[6]Please see Matlab documentation for fitensemble function.

[7]Or the dog in Image 8 in Figure 4.

[8]We obtained accuracy of 0.2399 ± 0.0016 (mean ± SD) over 60 runs of the RUSBoost classifier.

[9]Note that here we used task and cognitive state interchangeably. There are however subtle differences. Cognitive state refers to the state of a person's psychological condition (e.g., confusion, preoccupation, wonder, etc.). By task here referred to a well-defined question that observers should try to answer (e.g., estimating age, search for an object, reading, etc.).

[10]For prevalence statistics, visit http://www.cdc.gov/ncbddd/.

References

Albert, M., Toledo, S., Shapiro, M., & Kording, K. (2012). Using mobile phones for activity recognition in Parkinson's patients. *Frontiers in Neurology, 3,* 158.

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66–80.

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *Plos One, 8.*

Betz, T., Kietzmann, T., Wilming, N., & König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision, 10*(3):15, 1–14, http://www.journalofvision.org/content/10/3/15, doi:10.1167/10.3.15. [PubMed] [Article]

Borji, A. (2012). Boosting bottom-up and top-down visual features for saliency estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 438–445).

Borji, A., & Itti, L. (2013). State-of-the-art in modeling visual attention. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), 35,* 185–207.

Borji, A., Sihite, D. N., & Itti, L. (2012a). Probabilistic learning of task-specific visual attention. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 470–477).

Borji, A., Sihite, D. N., & Itti, L. (2012b). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Trans. Image Processing, 22,* 55–69.

Borji, A., Sihite, D. N., & Itti, L. (2014). What/where to look next? Modeling top-down visual attention in complex interactive environments. *IEEE Transactions on Systems, Man, and Cybernetics*, *Part A-Systems and Humans*, in press.

Borji, A., Tavakoli, H. R., Sihite, D. N., & Itti, L. (2013). Analysis of scores, datasets, and models in visual saliency prediction. *International Conference on Computer Vision (ICCV)* (pp. 921–928).

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9,* 27–38.

Bruce, N., & Tsotsos, J. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision, 9*(3):5, 1–24, http://www.journalofvision.org/content/9/3/5, doi:10.1167/9.3.5. [PubMed] [Article]

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33,* 741–753.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005).

A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291.

Buswell, G. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Coen-Cagli, R., Coraggio, P., Napoletano, P., Schwartz, O., Ferraro, M., & Boccignone, G. (2009). Visuomotor characterization of eye movements in a drawing task. *Vision Research, 49.*

Castelhano, M., Mack, M., & Henderson, J. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9*(3):6, 1–15, http://www.journalofvision.org/content/9/3/6, doi:10.1167/9.3.6. [PubMed] [Article]

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision, 9*(12):10, 1–15, http://www.journalofvision.org/content/9/12/10, doi:10.1167/9.12.10. [PubMed] [Article]

Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., & Wood, M. J. (2013). Temporal characteristics of overt attentional behaviour during category learning. *Attention Perception & Psychophysics, 75*(2), 244–256.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences, USA, 102,* 12629–12633.

Clark, J. J., & O'Regan, J. K. (1998). Word ambiguity and the optimal viewing position in reading. *Vision Research, 39,* 843–857.

Crespi, S., Robino, C., Silva, O., & deSperati, C. (2012). Spotting expertise in the eyes: Billiards knowledge as revealed by gaze shifts in a dynamic visual prediction task. *Journal of Vision, 12*(11):30, 1–19, http://www.journalofvision.org/content/12/11/30, doi:10.1167/12.11.30. [PubMed] [Article]

Cyganek, B., & Gruszczynski, S. (2014). Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. *Neurocomputing, 126,* 78–94.

DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition, 17,* 790–811.

Doshi, A., & Trivedi, M. M. (2009). On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *Intelligent Transportation Systems, IEEE Transactions on, 10,* 453–462.

Doshi, A., & Trivedi, M. M. (2012). Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of Vision, 12*(2):9, 1–16, http://www.journalofvision.org/content/12/2/9, doi:10.1167/12.2.9. [PubMed] [Article]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269–297.

Epelboim, J., & Suppes, P. (2001). A model of eye movements and visual working memory during problem solving in geometry. *Vision Research, 41,* 1561–1574.

Fathi, A., Farhadi, A., & Rehg, J. M. (2011). Understanding egocentric activities. *International Conference on Computer Vision (ICCV)* (pp. 407–414).

Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition, 119.*

Fix, E., & Hodges, J. L. (1951). *Discriminatory analysis, nonparametric discrimination: Consistency properties* (Technical Report No. 4). Randolph Field, TX: USAF School of Aviation Medicine.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 847.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 1,* 119–139.

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., & Dosil, R. (2012). Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image and Vision Computing, 30,* 51–64.

Greene, M., Liu, T., & Wolfe, J. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research, 62,* 1–8.

Griffiths, A. N., Marshall, R. W., & Richens, A. (1984). Saccadic eye movement analysis as a measure of drug effects on human psychomotor performance. *British Journal of Clinical Pharmacology, 18,* 73S–82S.

Guo, C., & Zhang, L. (2010). A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *IEEE Trans, on Image Processing, 19*(1), 185–198.

Hagemann, N., Schorer, J., Canal-Bruland, R., Lotz,

S., & Strauss, B. (2010). Visual perception in fencing: Do the eye movements of fencers represent their information pickup? *Attention, Perception & Psychophysics, 72,* 2204–2214.

Haji-Abolhassani, A., & Clark, J. J. (2013). A computational model for task inference in visual search. *Journal of Vision, 13*(3):29, 1–24, http://www.journalofvision.org/content/13/3/29, doi:10.1167/13.3.29. [PubMed] [Article]

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. *Advances in Neural Information Processing Systems (NIPS), 19,* 545–552.

Harel, J., Moran, C., Huth, A., Einhaeuser, W., & Koch, C. (2008). Decoding what people see from where they look: Predicting visual stimuli from scanpaths. In *Attention in Cognitive Systems* (pp. 15–26). Berlin: Springer.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9,* 188–194.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1):6, 49–63, http://www.journalofvision.org/content/3/1/6, doi:10.1167/3.1.6. [PubMed] [Article]

Henderson, J., Shinkareva, S., Wang, J., Luke, S., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *Plos One, 8*(5), e64937.

Hou, X., & Zhang, L. (2007). Saliency detection: A spectral residual approach. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–8).

Hou, X., & Zhang, L. (2008). Dynamic visual attention: Searching for coding length increments. *Advances in Neural Information Processing Systems (NIPS),* 681–688.

Iqbal, S., & Bailey, B. (2004). Using eye gaze patterns to identify user tasks. *Proceedings of the Grace Hopper Celebration of Women in Computing,* October 6–9, 2004, Chicago, IL, USA.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2,* 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Jang, Y.-M., Lee, S., Mallipeddi, R., Kwak, H.-W., & Lee, M. (2011). Recognition of human's implicit intention based on an eyeball movement pattern analysis. In B.-L. Lu, L. Zhang, & J. T. Kwok (Eds.), *Neural Information Processing* (pp. 138–145). Berlin: Springer.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20,* 146–154.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature.* E-pub ahead of print, http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12715.html?WT.mc_id=TWT_NatureNeuro, doi:10.1038/nature12715.

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). Learning to predict where humans look. *International Conference on Computer Vision (ICCV)* (pp. 2106–2113).

Kaakinen, J. K., & Hyönä, J. (2010). Task effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1561–1566.

Kanan, C., Ray, N., Bseiso, D. N. F., Hsiao, J. H., & Cottrell, G. W. (2014). Predicting an observer's task using multi-fixation pattern analysis. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA-2014).*

Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology, 19*(1), 52–56.

Klin, A., Lin, D., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature, 459*(7244), 257–261.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kuhn, G., & Tatler, B. W. (2005). Magic and fixation: Now you don't see it, now you do. *Perception, 34,* 1155–1161.

Kuhn, G., Tatler, B. W., Findlay, J. M., & Cole, G. G. (2008). Misdirection in magic: Implications for the relationship between eye gaze and attention. *Visual Cognition, 16*(2–3), 391–405.

Land, M. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research, 25,* 296–324.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3565.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369,* 742–744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3,* 1340–1345.

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology, 11,* 1215–1220.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28,* 1311–1328.

Lethaus, F., Baumann, M. R. K., Köster, F., & Lemmer, K. (2013). A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data. *Neurocomputing, 121,* 108–130.

Loetscher, T., Bockisch, C., Nicholls, M., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology, 20,* 264–265.

Macknik, S. L., King, M., Randi, J., Robbins, A., Teller, Thompson, J., & Martinez-Conde, S. (2008). Attention and awareness in stage magic: Turning tricks into research. *Nature Reviews Neuroscience, 9,* 871–879.

Mast, F., & Kosslyn, S. (2002). Eye movements during visual mental imagery. *Trends in Cognitive Sciences, 6*(7), 271–272.

Meijering, B., van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2012). What eye movements can tell about theory of mind in a strategic game. *Plos One, 7*(9), e45961.

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research, 179,* 427–442.

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision, 11*(8):17, 1–15, http://www.journalofvision.org/content/11/8/17, doi:10.1167/11.8.17. [PubMed] [Article]

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231.

O'Connell, T., & Walther, D. (2012). Fixation patterns predict scene category. *Journal of Vision, 12*(9): 801, http://www.journalofvision.org/content/12/9/801, doi:10.1167/12.9.801. [Abstract]

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

Peters, R., & Itti, L. (2007). Congruence between model and human attention reveals unique signatures of critical visual events. In *Advances in Neural Information Processing Systems (NIPS)*.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45,* 2397–2416.

Pirsiavash, H., & Ramanan, D. (2012). Detecting activities of daily living in first-person camera views. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2847–2854).

Poynter, W., Barber, M., Inman, J., & Wiggins, C. (2013). Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. *Vision Research, 89,* 32–38.

Rajashekar, J., Bovik, L., & Cormack, A. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, 6*(4):7, 379–386, http://www.journalofvision.org/content/6/4/7, doi:10.1167/6.4.7. [PubMed] [Article]

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception, 8,* 21–30.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*(1), 1–41.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The e-z reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–476.

Renninger, L. W., Coughlan, J. M., Verghese, P., & Malik, J. (2004). An information maximization model of eye movements. In *Advances in Neural Information Processing Systems (NIPS)*.

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.

Salzberg, S. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery, 1,* 317–328.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning, 2,* 197–227.

Schütz, A., Braun, D., & Gegenfurtner, K. (2011). Eye movements and perception: A selective review. *Journal of Vision, 11*(5):9, 1–30, http://www.journalofvision.org/content/11/5/9, doi:10.1167/11.5.9. [PubMed] [Article]

Seiffert, C., Khoshgoftaar, T. M., Van Hülse, J., & Napolitano, A. (2010). RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 40,* 185–197.

Seo, H., & Milanfar, P. (2009). Static and space-time visual saliency detection by self-resemblance. *Journal of Vision, 9*(12):15, 1–27, http://www.

journalofvision.org/content/9/12/15, doi:10.1167/9.12.15. [PubMed] [Article]

Shaffer, J. (1995). Multiple hypothesis testing. *Annual Review of Psych, 46,* 561–584.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74,* 1.

Sperling, G., & Dosher, B. A. (1986). *Strategy and optimization in human information processing*. University Park, PA: Citeseer.

Subramanian, R., Sebe, N., Kankanhalli, M., & Chua, T. (2010). An eye fixation database for saliency detection in images. In S. Ramanathan, H. Katti, N. Sebe, M. Kankanhali, & T.-S. Chua (Eds.), *Computer Vision—ECCV 2010* (pp. 30–43). Berlin: Springer.

Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, http://www.journalofvision.org/content/11/5/5, doi:10.1167/11.5.5. [PubMed] [Article]

Tatler, B., & Vincent, B. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17,* 1029–1054.

Tatler, B., Wade, N., Kwan, H., Findlay, J., & Velichkovsky, B. (2010). Yarbus, eye movements, and vision. *i-Perception, 1,* 7–27.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Tseng, P., Cameron, I. G. M., Pari, G., Reynolds, J. N., Munoz, D. P., & Itti, L. (2012). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology, 260*(1), 275–284.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8,* 167–190.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception, 36,* 91–100.

Windau, J., & Itti, L. (2013). Situation awareness via sensor-equipped eyeglasses. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5674–5679). IEEE.

Woods, D. J., Beecher, G. P., & Ris, M. D. (1978). The effects of stressful arousal on conjugate lateral eye movement. *Motivation and Emotion, 2,* 345–353.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and Performance, 18,* 73–103.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Zelinsky, G., Peng, Y., & Samaras, D. (2013). Eye can read your mind: Decoding gaze fixations to reveal categorical search targets. *Journal of Vision, 13*(14):10, 1–13, http://www.journalofvision.org/content/13/14/10, doi:10.1167/13.14.10. [PubMed] [Article]

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115,* 787–835.

Zelinsky, G., Zhang, W., & Samaras, D. (2008). Eye can read your mind: Decoding eye movements to reveal the targets of categorical search tasks. *Journal of Vision, 8*(6):380, http://www.journalofvision.org/content/8/6/380, doi:10.1167/8.6.380. [Abstract]

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). Sun: A Bayesian framework for saliency using natural statistics. *Journal of Vision, 8*(7):32, 1–20, http://www.journalofvision.org/content/8/7/32, doi:10.1167/8.7.32. [PubMed] [Article]

Zhao, Q., & Koch, C. (2012). Learning visual saliency by combining feature maps in a nonlinear manner using adaboost. *Journal of Vision, 12*(6):22, 1–15, http://www.journalofvision.org/content/12/6/22, doi:10.1167/12.6.22. [PubMed] [Article]

Appendix 1

Task decoding accuracies over single images in Experiment 1 using RUSBoost classifier. Please see Table 3.

Appendix 2

Task decoding accuracies over single images in Experiment 2 using RUSBoost classifier. Please see Table 4.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R1	0.5294	0.2941	0.4706	0.2353	0.2353	0.3529	0.2941	0.5294	0.4118	0.2941	0.4118	0.4118	0.1765	0.3529	0.3529	0.3529	0.1765	0.3529	0.1765	0.3529
R2	0.6471	0.4706	0.3529	0.2941	0.4118	0.2941	0.2941	0.5294	0.5294	0.2941	0.4118	0.4118	0.2353	0.4118	0.3529	0.4118	0.3529	0.3529	0.1176	0.3529
R3	0.4706	0.3529	0.2941	0.2353	0.3529	0.3529	0.2941	0.5294	0.4706	0.2353	0.4706	0.3529	0.1176	0.2941	0.3529	0.4118	0.2941	0.2941	0.1765	0.3529
R4	0.4118	0.2941	0.2353	0.2941	0.2353	0.3529	0.2353	0.4706	0.4706	0.2353	0.4118	0.2941	0.1176	0.4118	0.3529	0.3529	0.1765	0.3529	0.1176	0.3529
R5	0.4706	0.3529	0.4118	0.2941	0.2941	0.3529	0.1765	0.5294	0.5294	0.2941	0.4118	0.4118	0.1176	0.4118	0.3529	0.3529	0.1765	0.3529	0.1765	0.4118
R6	0.4706	0.4118	0.3529	0.2941	0.4118	0.3529	0.2353	0.4706	0.4706	0.2353	0.4706	0.4118	0.2353	0.3529	0.3529	0.3529	0.2941	0.2941	0.1176	0.3529
R7	0.6471	0.2941	0.3529	0.2353	0.2353	0.2941	0.1765	0.5294	0.4706	0.2941	0.4706	0.4118	0.2353	0.3529	0.3529	0.3529	0.1765	0.4118	0.2353	0.4118
R8	0.5294	0.3529	0.3529	0.2353	0.3529	0.3529	0.2353	0.5882	0.4706	0.3529	0.4706	0.4118	0.2941	0.2941	0.3529	0.3529	0.1765	0.3529	0.0588	0.3529
R9	0.4706	0.4118	0.4118	0.2353	0.3529	0.3529	0.2941	0.5294	0.4706	0.3529	0.3529	0.4118	0.1765	0.3529	0.3529	0.3529	0.2353	0.2941	0.1765	0.4118
R10	0.4706	0.4118	0.4118	0.2941	0.2941	0.3529	0.2353	0.4706	0.5294	0.2941	0.3529	0.5294	0.1176	0.4706	0.2941	0.4118	0.2353	0.3529	0.1765	0.3529
R11	0.5294	0.4118	0.2941	0.2941	0.2941	0.2941	0.2941	0.5882	0.4706	0.2353	0.4118	0.4118	0.1176	0.3529	0.4118	0.3529	0.1765	0.2941	0.0588	0.4118
R12	0.4118	0.2941	0.2353	0.2353	0.3529	0.4118	0.1765	0.5294	0.4118	0.2941	0.4118	0.4706	0.2353	0.2941	0.4706	0.3529	0.2353	0.3529	0.1176	0.3529
R13	0.4118	0.4118	0.3529	0.2941	0.4118	0.4706	0.2941	0.5294	0.4706	0.2941	0.3529	0.2941	0.1765	0.4118	0.4706	0.2941	0.2353	0.2941	0	0.3529
R14	0.5882	0.3529	0.3529	0.2941	0.2941	0.4118	0.2353	0.5294	0.4706	0.2941	0.4118	0.5294	0.1176	0.2941	0.3529	0.3529	0.2941	0.2353	0.1765	0.3529
R15	0.5294	0.4118	0.3529	0.3529	0.2353	0.4118	0.2941	0.4706	0.5294	0.2941	0.4118	0.4118	0.2353	0.3529	0.4118	0.3529	0.1765	0.2941	0.0588	0.3529
R16	0.4118	0.4118	0.2941	0.2353	0.2353	0.4118	0.3529	0.5294	0.4706	0.2353	0.4118	0.3529	0.1765	0.4118	0.4118	0.4118	0.1765	0.4118	0.1765	0.4118
R17	0.4706	0.4118	0.3529	0.2941	0.2353	0.3529	0.3529	0.4706	0.4706	0.2353	0.3529	0.4118	0.1765	0.3529	0.4118	0.3529	0.1765	0.2941	0.0588	0.3529
R18	0.4706	0.2941	0.4118	0.2353	0.2941	0.3529	0.2353	0.4706	0.5294	0.2941	0.4118	0.4706	0.1765	0.2353	0.3529	0.3529	0.2353	0.2941	0.2353	0.3529
R19	0.4118	0.3529	0.2941	0.3529	0.2941	0.2941	0.2353	0.5294	0.4706	0.2353	0.2941	0.4118	0.1176	0.3529	0.3529	0.3529	0.2353	0.3529	0.1176	0.3529
R20	0.4118	0.4118	0.4118	0.2353	0.3529	0.4118	0.2353	0.5294	0.4706	0.2941	0.4118	0.3529	0.2353	0.3529	0.4118	0.3529	0.1765	0.2941	0.1176	0.4118
R21	0.4706	0.4706	0.2941	0.3529	0.2353	0.3529	0.2941	0.5294	0.4706	0.2353	0.4118	0.4118	0.2941	0.2941	0.2941	0.3529	0.2941	0.2353	0.0588	0.3529
R22	0.4706	0.2941	0.2941	0.2353	0.3529	0.3529	0.1765	0.4706	0.4706	0.2941	0.4118	0.4706	0.2353	0.3529	0.4118	0.3529	0.2941	0.2353	0.1765	0.3529
R23	0.5882	0.2941	0.2941	0.4118	0.3529	0.3529	0.2941	0.5294	0.4706	0.2941	0.4118	0.4118	0.1765	0.2941	0.2941	0.3529	0.2941	0.2941	0.0588	0.3529
R24	0.5294	0.3529	0.3529	0.2353	0.3529	0.2941	0.2941	0.5882	0.5294	0.2941	0.4118	0.4118	0.1765	0.4118	0.2941	0.2941	0.1765	0.4118	0.1765	0.4118
R25	0.4706	0.2941	0.2941	0.1765	0.3529	0.3529	0.3529	0.4706	0.4706	0.2353	0.4118	0.3529	0.2941	0.4118	0.3529	0.3529	0.2353	0.3529	0.0588	0.3529
R26	0.5294	0.5294	0.2941	0.3529	0.2941	0.2941	0.2353	0.5294	0.4706	0.2353	0.3529	0.3529	0.1765	0.4118	0.4706	0.3529	0.1765	0.4118	0.1176	0.3529
R27	0.4118	0.3529	0.3529	0.2941	0.2353	0.4118	0.2941	0.5294	0.4706	0.2941	0.4118	0.3529	0.1765	0.3529	0.4118	0.3529	0.2353	0.3529	0.1176	0.3529
R28	0.4706	0.4118	0.4118	0.2353	0.4118	0.2941	0.2353	0.5294	0.5294	0.2353	0.4706	0.3529	0.1765	0.4118	0.3529	0.3529	0.2353	0.2353	0.1176	0.3529
R29	0.5294	0.4118	0.2353	0.2353	0.2941	0.4118	0.2941	0.5294	0.4706	0.2941	0.3529	0.4118	0.1765	0.2941	0.4706	0.3529	0.1765	0.3529	0.1176	0.3529
R30	0.5294	0.3529	0.4118	0.4118	0.2941	0.3529	0.2353	0.4706	0.4706	0.2353	0.4118	0.4118	0.1765	0.5294	0.4118	0.3529	0.1765	0.4118	0.1176	0.3529
R31	0.4706	0.2941	0.2941	0.2353	0.2941	0.2941	0.3529	0.5294	0.4706	0.2353	0.3529	0.3529	0.1765	0.4118	0.4118	0.3529	0.2353	0.4706	0.1176	0.3529
R32	0.5294	0.3529	0.3529	0.2941	0.2941	0.2941	0.2941	0.5294	0.4706	0.2941	0.4706	0.4118	0.1765	0.4118	0.4118	0.3529	0.1765	0.2941	0.1765	0.4118
R33	0.4706	0.4706	0.4118	0.2941	0.2941	0.3529	0.2353	0.5294	0.4706	0.3529	0.4706	0.4118	0.1176	0.3529	0.2941	0.4118	0.2353	0.2941	0.0588	0.4118
R34	0.5294	0.4118	0.2941	0.2353	0.2941	0.4118	0.2941	0.5294	0.5294	0.2941	0.4118	0.4706	0.1176	0.4118	0.3529	0.4118	0.1765	0.4118	0.1176	0.4118
R35	0.4118	0.4706	0.2941	0.2941	0.2941	0.3529	0.1176	0.4706	0.4706	0.2941	0.4118	0.3529	0.1765	0.4706	0.4118	0.3529	0.1765	0.2941	0.1765	0.3529
R36	0.5294	0.3529	0.3529	0.2941	0.2941	0.2941	0.2941	0.5294	0.5882	0.2941	0.4118	0.4118	0.1765	0.3529	0.2941	0.2941	0.1765	0.4118	0.0588	0.3529
R37	0.4706	0.3529	0.3529	0.2941	0.2353	0.4118	0.1765	0.4706	0.4706	0.1765	0.2941	0.4118	0.1176	0.3529	0.4118	0.3529	0.2353	0.2941	0.1176	0.4118
R38	0.5882	0.3529	0.3529	0.2941	0.3529	0.3529	0.3529	0.4706	0.4706	0.3529	0.4706	0.3529	0.2353	0.4118	0.4118	0.3529	0.1765	0.3529	0.1176	0.3529
R39	0.4706	0.3529	0.4706	0.3529	0.2941	0.3529	0.2353	0.4706	0.5294	0.2941	0.4118	0.4118	0.2353	0.3529	0.3529	0.3529	0.2353	0.2941	0.2353	0.3529
R40	0.5294	0.4118	0.3529	0.2941	0.2941	0.3529	0.2941	0.4706	0.4706	0.3529	0.4706	0.3529	0.2353	0.4118	0.4706	0.3529	0.1765	0.2941	0.1765	0.3529
R41	0.4706	0.4706	0.4118	0.2353	0.2941	0.2941	0.2941	0.5294	0.5294	0.2941	0.3529	0.4706	0.2353	0.2353	0.4706	0.3529	0.2353	0.3529	0.1176	0.3529
R42	0.4706	0.2941	0.3529	0.2941	0.2941	0.3529	0.2353	0.5294	0.4118	0.3529	0.2353	0.4118	0.1765	0.4706	0.3529	0.3529	0.2353	0.3529	0.0588	0.3529
R43	0.4118	0.2941	0.4118	0.2941	0.2941	0.3529	0.3529	0.4706	0.4706	0.2941	0.3529	0.3529	0.2353	0.3529	0.3529	0.3529	0.2941	0.2941	0.1765	0.3529
R44	0.4706	0.5294	0.2941	0.2353	0.2353	0.3529	0.2941	0.4706	0.5294	0.2353	0.2941	0.3529	0.1765	0.3529	0.2941	0.3529	0.2941	0.2941	0.1765	0.4118
R45	0.4706	0.2941	0.2941	0.2941	0.4118	0.3529	0.2941	0.5294	0.4118	0.2941	0.4118	0.2941	0.1765	0.3529	0.4118	0.3529	0.1765	0.4118	0.1176	0.3529
R46	0.5294	0.3529	0.3529	0.3529	0.2353	0.4118	0.3529	0.4706	0.4706	0.2941	0.3529	0.3529	0.2353	0.3529	0.4118	0.5294	0.3529	0.2941	0.1176	0.3529
R47	0.5294	0.2941	0.2941	0.2941	0.3529	0.3529	0.2353	0.4706	0.5294	0.2941	0.4118	0.3529	0.1176	0.2353	0.4118	0.3529	0.2353	0.2353	0.1765	0.3529
R48	0.4118	0.3529	0.4118	0.2941	0.2941	0.2941	0.2941	0.4706	0.4706	0.2353	0.4118	0.3529	0.1765	0.4118	0.3529	0.3529	0.1765	0.2941	0.1176	0.3529
R49	0.3529	0.3529	0.3529	0.2353	0.3529	0.4118	0.2941	0.5294	0.5294	0.2941	0.4706	0.3529	0.1176	0.3529	0.2941	0.3529	0.2353	0.2353	0.1176	0.4118
R50	0.5294	0.4118	0.4118	0.3529	0.3529	0.3529	0.2941	0.5294	0.4118	0.2353	0.4118	0.3529	0.2941	0.2941	0.3529	0.3529	0.2353	0.2941	0.1176	0.3529
Avg0.	0.4906	0.3741	0.3518	0.2812	0.3094	0.3565	0.2694	**0.5118**	0.4835	0.2800	0.3977	0.3965	0.1859	0.3623	0.3765	0.3600	0.2259	0.3270	**0.1318**	0.3670
p-val0.	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0152	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0011	0.0000	0.0000	0.0000

Table 3. Performance of the RUSBoost classifier for task decoding per image in Experiment 1 using Feature Type 3. Columns represent Images 1 to 20 and each row corresponds to a separate run. Each single number is the average of 17 accuracies (i.e., leave one subject out). Last row shows p values across RUSBoost runs using t test (vs. chance). Easiest and most difficult stimuli are shown in bold-face font.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R1	0.1905	0.3333	0.3333	0.2381	0.2381	0.3333	0.1429	0.1905	0.2857	0.4286	0.1905	0.3810	0.2381	0.2857	0.2857
R2	0.1905	0.3333	0.2381	0.3333	0.1905	0.2857	0.0476	0.2857	0.2857	0.3810	0.1429	0.3810	0.2857	0.3810	0.2381
R3	0.1905	0.3333	0.3810	0.2381	0.2381	0.2857	0.0952	0.2381	0.3333	0.3810	0.1429	0.4286	0.2381	0.3333	0.2381
R4	0.0952	0.2857	0.3333	0.1905	0.2381	0.1905	0.0952	0.2381	0.2381	0.4286	0.1429	0.3333	0.3810	0.3810	0.2381
R5	0.1905	0.3810	0.3810	0.2381	0.1905	0.2381	0.1429	0.2381	0.2857	0.4762	0.2381	0.3810	0.3333	0.2857	0.2381
R6	0.1429	0.4762	0.3333	0.2381	0.2381	0.1905	0.0952	0.1905	0.2857	0.3810	0.0952	0.3810	0.2857	0.3333	0.3333
R7	0.0952	0.3333	0.3333	0.2381	0.3333	0.2381	0.0952	0.2381	0.2857	0.3810	0.2381	0.4286	0.2857	0.3333	0.2381
R8	0.1905	0.3333	0.3333	0.2381	0.2381	0.3333	0.1905	0.3333	0.2381	0.4286	0.2857	0.3810	0.2857	0.3810	0.2857
R9	0.1429	0.3333	0.2857	0.2857	0.2381	0.2857	0.1429	0.2857	0.2857	0.4286	0.1429	0.3333	0.2857	0.3810	0.2381
R10	0.0952	0.4286	0.3333	0.1905	0.2857	0.2381	0.0952	0.2381	0.2857	0.3810	0.1429	0.4286	0.2381	0.4286	0.2381
R11	0.1905	0.3810	0.2857	0.2381	0.2381	0.2381	0.1429	0.2381	0.2381	0.4286	0.1905	0.4286	0.2381	0.3810	0.1905
R12	0.0952	0.3333	0.3333	0.1429	0.2857	0.3333	0	0.2857	0.2857	0.4286	0.1429	0.3810	0.2381	0.2857	0.2381
R13	0.1429	0.3810	0.3810	0.2381	0.2381	0.2857	0.1429	0.3333	0.2381	0.3810	0.1429	0.3810	0.2857	0.3810	0.2381
R14	0.2381	0.4762	0.3810	0.2381	0.2381	0.2381	0.0476	0.2857	0.2381	0.4286	0.0952	0.3333	0.2381	0.3810	0.2857
R15	0.2381	0.3810	0.2381	0.2857	0.2381	0.1905	0.1429	0.2857	0.2381	0.4286	0.1429	0.3333	0.3333	0.2381	0.2857
R16	0.1905	0.3810	0.2857	0.1905	0.3810	0.2381	0.0476	0.2381	0.2857	0.4762	0.1429	0.3333	0.1905	0.3333	0.2857
R17	0.1905	0.4286	0.2857	0.1905	0.2857	0.2381	0.0476	0.2857	0.2857	0.3810	0.1429	0.3810	0.2381	0.4286	0.2857
R18	0.1429	0.3333	0.3333	0.1905	0.2857	0.2857	0.0952	0.3333	0.2857	0.4762	0.1429	0.3810	0.2381	0.4286	0.2381
R19	0.1429	0.4286	0.2381	0.2381	0.2381	0.2857	0.1429	0.2857	0.2857	0.3810	0.1429	0.4286	0.2857	0.3810	0.1905
R20	0.0952	0.3810	0.3810	0.2381	0.2381	0.3333	0.1429	0.2857	0.2857	0.4762	0.1429	0.4286	0.2857	0.2857	0.2381
R21	0.1429	0.3333	0.1905	0.1905	0.2857	0.2857	0.1429	0.3333	0.2857	0.4286	0.1429	0.3810	0.1905	0.3810	0.2381
R22	0.1905	0.3333	0.2857	0.1905	0.2381	0.3333	0.0952	0.2857	0.2381	0.3810	0.0952	0.4286	0.2381	0.3810	0.2381
R23	0.1429	0.3333	0.3810	0.2381	0.1905	0.3333	0.0952	0.3333	0.4286	0.4286	0.0476	0.4286	0.2857	0.2857	0.2381
R24	0.1905	0.4286	0.2857	0.2857	0.2381	0.2381	0.1429	0.2857	0.3333	0.4286	0.1905	0.3810	0.2381	0.2857	0.2857
R25	0.1429	0.4286	0.2857	0.2857	0.2857	0.2857	0.0476	0.2857	0.1905	0.3810	0.0952	0.3810	0.2857	0.3333	0.2381
R26	0.1429	0.3333	0.3333	0.2381	0.3333	0.3333	0.0952	0.2381	0.2857	0.4286	0.0952	0.3810	0.2381	0.3810	0.2381
R27	0.2857	0.3333	0.3333	0.2857	0.2857	0.3333	0.1905	0.2857	0.3333	0.4286	0.2381	0.3810	0.2381	0.3333	0.2857
R28	0.1429	0.3810	0.2381	0.1905	0.2857	0.3333	0.1905	0.2857	0.3333	0.3810	0.0952	0.3810	0.2857	0.3810	0.1905
R29	0.1429	0.3333	0.4286	0.2857	0.2381	0.3810	0.0476	0.1905	0.2857	0.4286	0.0476	0.4286	0.2857	0.4286	0.2381
R30	0.1429	0.4286	0.3333	0.2381	0.2857	0.2857	0.1429	0.2857	0.3810	0.4762	0.1905	0.3810	0.2857	0.4286	0.3333
R31	0.1429	0.4286	0.3810	0.1905	0.2381	0.1905	0.1429	0.2857	0.2381	0.3810	0.0952	0.3810	0.2857	0.3333	0.1905
R32	0.1429	0.3810	0.2857	0.2857	0.2857	0.1905	0.0952	0.2381	0.2857	0.3810	0.1905	0.4286	0.2381	0.3810	0.2857
R33	0.1429	0.3333	0.3333	0.3333	0.2381	0.3333	0.1429	0.2381	0.2857	0.4286	0.2857	0.4286	0.1905	0.3810	0.2381
R34	0.1429	0.4286	0.2857	0.2381	0.2857	0.2857	0.0476	0.1905	0.2857	0.4286	0.1429	0.4286	0.2381	0.3810	0.2381
R35	0.1905	0.3333	0.3810	0.2381	0.2381	0.2857	0.1429	0.2381	0.2857	0.4762	0.1905	0.3333	0.2381	0.4286	0.1905
R36	0.1429	0.3810	0.3333	0.2381	0.2857	0.3333	0.0952	0.2381	0.2857	0.4762	0.1905	0.3810	0.1905	0.4286	0.2381
R37	0.1429	0.3333	0.3333	0.1905	0.2857	0.2381	0.0476	0.3333	0.3333	0.4286	0.2381	0.3810	0.2381	0.3810	0.3333
R38	0.1905	0.3810	0.3810	0.2381	0.3333	0.2857	0.0476	0.2857	0.2381	0.3333	0.0952	0.4286	0.2857	0.4286	0.1905
R39	0.1429	0.3810	0.3333	0.1429	0.2381	0.3333	0.0952	0.2381	0.2857	0.3810	0.2381	0.3810	0.2381	0.3333	0.2381
R40	0.1905	0.2857	0.2381	0.1429	0.2857	0.3333	0.0476	0.2857	0.3333	0.3810	0.0952	0.4286	0.2381	0.4286	0.2857
R41	0.1429	0.3333	0.3333	0.2857	0.2381	0.3333	0.0476	0.2381	0.3333	0.3810	0.1429	0.3810	0.1905	0.4286	0.1905
R42	0.1905	0.3333	0.2381	0.3333	0.1905	0.2857	0.0476	0.2857	0.2857	0.3810	0.1429	0.3810	0.2857	0.3810	0.2381
R43	0.1429	0.3810	0.3810	0.2381	0.2381	0.2857	0.1429	0.3333	0.2381	0.3810	0.1429	0.3810	0.2857	0.3810	0.2381
R44	0.2381	0.4762	0.3810	0.2381	0.2381	0.2381	0.0476	0.2857	0.2381	0.4286	0.0952	0.3333	0.2381	0.3810	0.2857
R45	0.2381	0.3810	0.2381	0.2857	0.2381	0.1905	0.1429	0.2857	0.2381	0.4286	0.1429	0.3333	0.3333	0.2381	0.2857
R46	0.1905	0.3810	0.2857	0.1905	0.3810	0.2381	0.0476	0.2381	0.2857	0.4762	0.1429	0.3333	0.1905	0.3333	0.2857
R47	0.1905	0.4286	0.2857	0.1905	0.2857	0.2381	0.0476	0.2857	0.2857	0.3810	0.1429	0.3810	0.2381	0.4286	0.2857
R48	0.1429	0.3333	0.3333	0.1905	0.2857	0.2857	0.0952	0.3333	0.2857	0.4762	0.1429	0.3810	0.2381	0.4286	0.2381
R49	0.1429	0.4286	0.2381	0.2381	0.2381	0.2857	0.1429	0.2857	0.2857	0.3810	0.1429	0.4286	0.2857	0.3810	0.1905
R50	0.1905	0.3810	0.2381	0.2857	0.2857	0.2381	0.1429	0.3333	0.2381	0.4286	0.1429	0.3810	0.3333	0.2857	0.2857
Avg.	0.1648	0.3733	0.3152	0.2352	0.2619	0.2771	**0.1019**	0.2724	0.2828	**0.4162**	0.1514	0.3867	0.2600	0.3648	0.2505
p-val.	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2602	0.0000	0.0000	0.0000	0.0000

Table 4. Performance of the RUSBoost classifier for task decoding in Experiment 2 using one observer out procedure. Columns represent Images 1 to 15 and each row corresponds to an individual run. Chance level is at 14.29%. Results are using the first two feature types (i.e., an 11610D vector). Shown by the last row *p* values (across RUSBoost runs), decoding is significantly above chance for some images, is significantly below chance for Image 7, and is nonsignificant versus chance for Image 11 (using *t* test).