

## Review

## Neural decoding of collective wisdom with multi-brain computing

Miguel P. Eckstein<sup>a,b,\*</sup>, Koel Das<sup>a,b</sup>, Binh T. Pham<sup>a</sup>, Matthew F. Peterson<sup>a</sup>, Craig K. Abbey<sup>a</sup>, Jocelyn L. Sy<sup>a</sup>, Barry Giesbrecht<sup>a,b</sup>

<sup>a</sup> Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, 93101, USA

<sup>b</sup> Institute for Collaborative Biotechnologies, University of California, Santa Barbara, Santa Barbara, CA, 93101, USA

## ARTICLE INFO

## Article history:

Received 7 April 2011

Revised 27 June 2011

Accepted 4 July 2011

Available online 14 July 2011

## Keywords:

Multi-brain computing

Collective wisdom

Neural decoding

Perceptual decisions

Group decisions

## ABSTRACT

Group decisions and even aggregation of multiple opinions lead to greater decision accuracy, a phenomenon known as collective wisdom. Little is known about the neural basis of collective wisdom and whether its benefits arise in late decision stages or in early sensory coding. Here, we use electroencephalography and multi-brain computing with twenty humans making perceptual decisions to show that combining neural activity across brains increases decision accuracy paralleling the improvements shown by aggregating the observers' opinions. Although the largest gains result from an optimal linear combination of neural decision variables across brains, a simpler neural majority decision rule, ubiquitous in human behavior, results in substantial benefits. In contrast, an extreme neural response rule, akin to a group following the most extreme opinion, results in the least improvement with group size. Analyses controlling for number of electrodes and time-points while increasing number of brains demonstrate unique benefits arising from integrating neural activity across different brains. The benefits of multi-brain integration are present in neural activity as early as 200 ms after stimulus presentation in lateral occipital sites and no additional benefits arise in decision related neural activity. Sensory-related neural activity can predict collective choices reached by aggregating individual opinions, voting results, and decision confidence as accurately as neural activity related to decision components. Estimation of the potential for the collective to execute fast decisions by combining information across numerous brains, a strategy prevalent in many animals, shows large time-savings. Together, the findings suggest that for perceptual decisions the neural activity supporting collective wisdom and decisions arises in early sensory stages and that many properties of collective cognition are explainable by the neural coding of information across multiple brains. Finally, our methods highlight the potential of multi-brain computing as a technique to rapidly and in parallel gather increased information about the environment as well as to access collective perceptual/cognitive choices and mental states.

© 2011 Elsevier Inc. All rights reserved.

## Contents

Introduction . . . . .	95
Neural vs. behavioral benefits of collective wisdom . . . . .	96
Relationship between multi-brain neural activity and behavioral group choices and decision confidence . . . . .	96
Early sensory vs. late decision related neural activity . . . . .	96
Decision time savings from integration of neural activity across brains. . . . .	96
Materials and methods . . . . .	96
Observers and procedure . . . . .	96
Stimuli and display . . . . .	96
Electroencephalogram data acquisition and pre-processing . . . . .	97
Feature reduction for each individual . . . . .	97
Pattern classifier for separate EEG windows of 20 time-points. . . . .	97
Combining EEG data across windows . . . . .	98
Peak N170 amplitude and mean amplitude from PO8 electrode . . . . .	98
Combining behavioral ratings and neural decision variables across individuals . . . . .	98
Optimal linear combination . . . . .	99

\* Corresponding author at: Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9663, USA. Fax: +1 805 893 4303.  
E-mail address: [eckstein@psych.ucsb.edu](mailto:eckstein@psych.ucsb.edu) (M.P. Eckstein).

Majority decision rule . . . . .	99
Extreme decision rule . . . . .	99
Prediction for performance improvement if observers were independent detectors. . . . .	99
Effect of correlation across observer responses on the relationship between performance of optimal linear combination and majority decision rules . . . . .	100
Cumulative integration of EEG data across time . . . . .	100
Theoretical calculation of classifier performance from cumulative integration of EEG data across time for the scenario of temporally independent samples . . . . .	100
Controls to ensure no over-fitting of the pattern classifier to the data . . . . .	100
Procedures to predict observer choices . . . . .	100
Evaluation of performance using Area under the ROC . . . . .	100
Evaluation of performance using proportion correct . . . . .	101
Estimation of error in measurements of pattern classifier performance on individual observer data . . . . .	101
Estimation of error in measurements of performance after combining multiple observer data. . . . .	101
Across brain vs. within brain pattern classifier performance . . . . .	101
Results . . . . .	101
Behavioral and pattern classifier performance of individual observers . . . . .	101
Pooling behavioral decisions across observers . . . . .	101
Pooling neural decision variables across multiple brains . . . . .	101
Unique benefits of integrating information across multiple brains . . . . .	102
Performance benefits from integrating neural decision variables in early sensory vs. late decision stages . . . . .	102
Choice probabilities of groups for early sensory vs. late decision neural activity . . . . .	102
Spatiotemporal distribution of correlations between group behavioral decision confidence/voting outcome and neural activity . . . . .	103
Incremental temporal integration of EEG activity and faster potential decisions when combining neural decision variables across brains . . . . .	103
Using neural activity to rank order groups of individuals based on the accuracy of their decisions . . . . .	105
Discussion . . . . .	105
Multi-brain computing and improved neural decoding of the state of the world . . . . .	105
Benefits of collective wisdom for perceptual decisions: sensory coding vs. late decision . . . . .	105
Relationship between collective neural activity and behavioral decision confidence . . . . .	106
Time-savings by integrating neural activity across brains . . . . .	106
Comparing group decision rules applied to neural decision variables and behavioral opinions. . . . .	107
Practical considerations, applications, and limitations of current study. . . . .	107
Acknowledgments . . . . .	108
Appendix A. Supplementary data. . . . .	108
References. . . . .	108

## Introduction

Many animals rely on groups to make decisions in tasks such as foraging for food, placement of their nests, and navigation (Cousin, 2009; Conradt and Roper, 2009; Conradt and List, 2009). Honeybee scouts dance around possible future homes for a nest site but then jointly converge towards one site (Seeley and Buhrman, 1999). Birds migrate in groups to increase navigational accuracy (Simons, 2004) and cockroaches collectively decide on shelter selection (Halloy et al., 2007). Collective decisions lead to more accurate and faster decisions and thus provide clear benefits for survival. Humans also typically benefit from making decisions in groups (Sorkin et al., 2001; Laughlin et al., 2002) and important decisions such as medical, judicial and governmental decisions are commonly made in consultation with a group of experts. These benefits result not only from human social interaction; even the aggregation of multiple human opinions through simple averaging, majority rules or more sophisticated algorithms can also lead to remarkable improvements in decision accuracy. This was first shown by Francis Galton who analyzed the opinions of 787 people about the weight of an ox and found that combining their numerical assessments resulted in a median estimate that was remarkably close to the true weight of the ox (Galton, 1907). The same principle is also manifested in present day futures markets (Surowiecki, 2005). Furthermore, modern human society is increasingly organized around the aggregation of collective opinions reflected in people's increased use of web ratings for daily choices about consumer products, lodging, food and entertainment.

Despite its ubiquity in the animal kingdom and modern human society, little is known about the neural basis of collective wisdom. The benefits of combining multiple opinions might arise from the

integration of noisy neural information about the impending choices represented in the brains of the participating individuals. This would suggest that there is more accurate information related to a decision represented in multiple brains than that in an individual brain because each brain might encode different aspects of a sensory event and also because of noise reduction through brain pooling. But can we obtain a neural measurement of the additional information stemming from the integration of information across multiple sensory and decision neural systems? At what neural processing stage does the benefit arise? Visual perceptual decisions follow a neural processing cascade starting with the early cortical visual areas in the occipital lobe and then through the dorsal and ventral streams to parietal and prefrontal areas which have been related to perceptual decision making and categorization (Heekeren et al., 2004; Heekeren et al., 2006). Do the benefits of collective wisdom arise in later decision stages or are they present even in early sensory coding stages? Furthermore, if the increased neural information about the stimulus encoded in multiple brains is related to the decision reached by aggregating human opinions, then we should be able to predict trial-to-trial group choices from neural activity (e.g., Parker and Newsome, 1998; Gold and Shadlen, 2007; for electroencephalography, EEG; Philiastides and Sajda, 2006a).

In addition, the possibility of using machine learning algorithms to fuse neural activity across multiple individuals might represent a starting point for a new field of multi-brain computing, akin to multi-sensor fusion in engineering (Luo and Kay, 1989; Shah et al., 1997), but which focuses on gathering more information through combining measurements across different biological sensors rather than physical detectors. The technology could be a contribution to the rising field of neuroergonomics (Parasuraman, 2003; Parasuraman and Rizzo, 2007).

and could be potentially used for improvement of neural-based object classification techniques, for prediction of group decisions from neural activity and monitoring collective cognitive and mental states.

Here, we applied neural decoding methods involving multivariate pattern classifiers (Parra et al., 2005; Kamitani and Tong, 2005; Haynes and Rees, 2005) to the EEG activity of twenty humans categorizing images of objects (faces vs. cars) in noise and introduce multi-brain computing methods to investigate the inherent information relevant to the perceptual categorization task in the brains of multiple human observers quantified using statistical decision theory. Our motivation is two-fold: to advance our understanding of the neural basis of collective wisdom by evaluating the relationship between integrated neural activity and group decisions based on behavioral opinions and also to evaluate the potential utility of multi-brain computing. Rather than have individuals interact to reach a group decision we obtained a group choice by applying simple rules to aggregate individual opinions expressed using a confidence rating.<sup>1</sup> For example, the simple majority rule (Sorkin et al., 1998) has been well-documented as a common decision rule used by humans (Hastie and Kameda, 2005) and can easily be mimicked by aggregating opinions and/or neural decision variables extracted from brain activity. The majority rule can be compared to other decision rules such as an optimal linear decision rule and/or an extreme opinion scheme. Following, we outline the specific theoretical and practical questions addressed in the current work.

#### *Neural vs. behavioral benefits of collective wisdom*

First, we evaluate whether the accuracy predicting the perceptual stimulus presented to the observers increases when integrating multi-brain activity in a similar manner to the performance benefits resulting from aggregating observers' behavioral opinions. We determine how many brains' EEGs need to be combined to achieve a collective neural-based classifier that achieves the behavioral accuracy discriminating the stimuli of a single human observer. We investigate how different combination rules compare when integrating neural decision variables across multiple brains. In particular, we assess how a neural implementation of the majority rule, common in humans and some animals (Couzin, 2009; Conrads and List, 2009; Conrads and Roper, 2009; Laughlin et al., 2002; Surowiecki, 2005), compares to the more resource demanding optimal linear pooling and a suboptimal neural extreme decision rule, and their relation to using similar decision rules to pool observers' behavioral opinions.

#### *Relationship between multi-brain neural activity and behavioral group choices and decision confidence*

To assess the relationship between the stimulus-related EEG and the choice reached by aggregating observer opinions, we evaluate: 1) the trial-to-trial correspondence (choice probability) between the collective neural-based classifier's decision and that achieved from aggregating the observers' opinions. 2) The correlation between voting results using a majority decision based on the behavioral opinions and those resulting from using the neural decision variables. 3) Whether there is a relationship between the collective neural decision variable and the group decision confidence determined by linearly pooling the observer opinions. 4) Finally, we assess the ability to rank order the behavioral performance of two random groups of individuals from their collective neural activity.

<sup>1</sup> Although there are clear limitations of not studying the dynamics of the group decisions, one advantage of obtaining group decisions from aggregation of individual opinions is that it allows for statistically intensive estimation of performance benefits across a larger number of group configurations and sizes that would otherwise be difficult to evaluate experimentally (i.e., group size > 10).

#### *Early sensory vs. late decision related neural activity*

To evaluate whether the benefits of collective integration of perceptual judgments arise in early sensory coding vs. late decision stages we take advantage of previous work establishing EEG time-epochs associated with early sensory activity (150–220 ms post stimulus onset; Philiastides and Sajda, 2006b; Philiastides and Heekeren, 2009) and a late decision component (300–400 ms after stimulus onset) for a similar car vs. face task. In particular, we evaluate neural pattern classifier performance for different temporal windows of EEG activity to assess whether the improvements in accuracy identifying the stimulus from neural activity arising from integration of neural decision variables across brains is different for the early sensory and late-decision time-epochs. In addition, we compare the ability of early sensory related EEG vs. late decision related activity to predict the choices reached by aggregating observers' behavioral opinions.

#### *Decision time savings from integration of neural activity across brains*

Studies have shown how collective decisions might allow groups of animals to more rapidly take actions such as in navigation of birds (Couzin, 2009). Such findings imply that analyses using neural activity across multiple brains rather than an individual brain should permit reaching desired decision accuracies at an earlier time. Yet, the relative benefits of integrating neural activity across time vs. across brains depend on the temporal correlation of the neural signals within individual brains as well as the correlation of neural activity across brains. Here, we estimate the potential time-savings in reaching a decision with a given accuracy from integrating neural activity across brains.

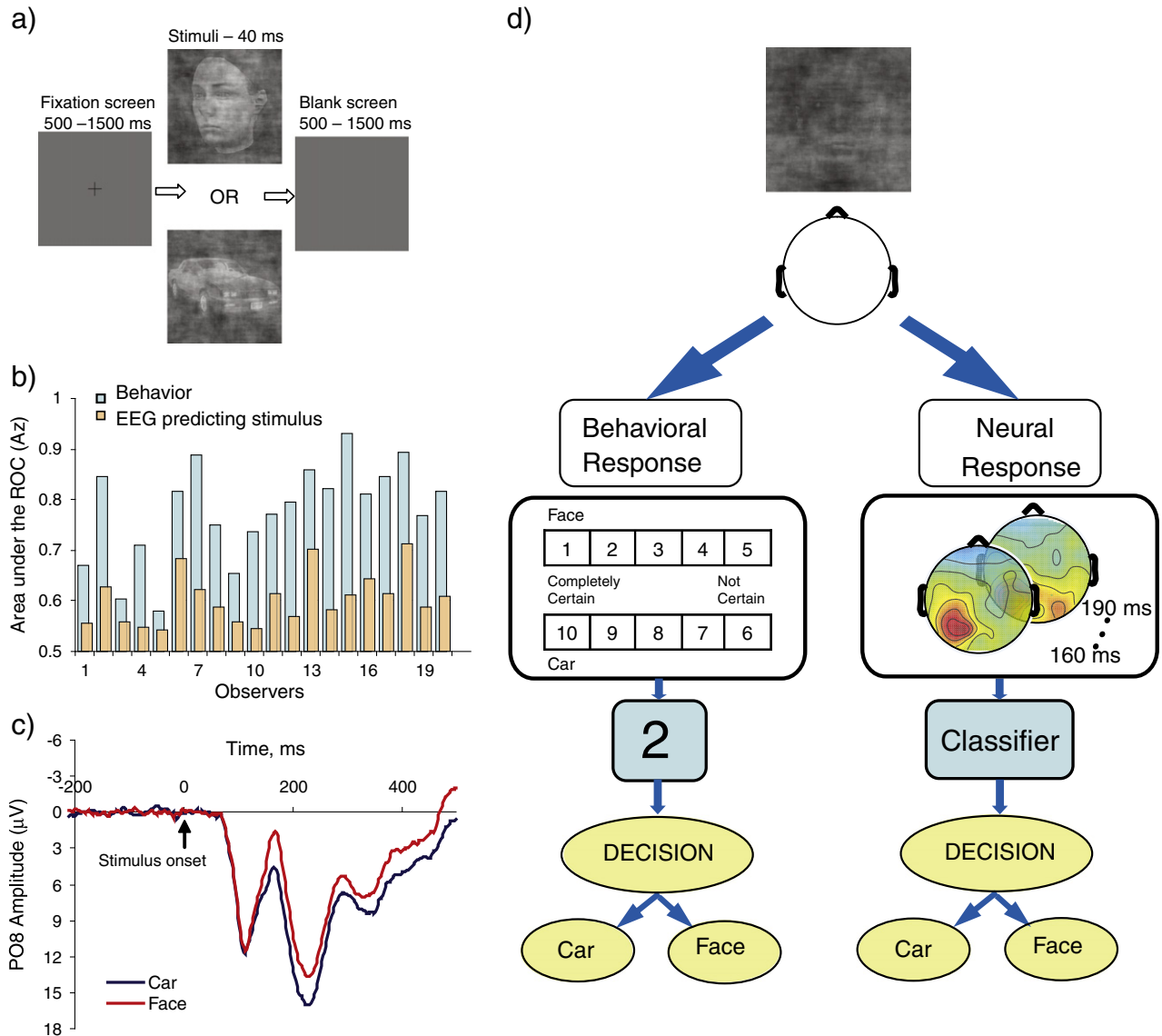
### **Materials and methods**

#### *Observers and procedure*

Twenty observers (ages: 18–26) naïve to the purpose of the study participated in the experiment. The study consisted of 1000 trials split into 5 successive sessions of 200 trials (preceded by 1000 stimulus-familiarization trials). Observers began a trial by fixating a central cross and clicking the mouse button. After a variable delay of 0.5–1.5 s the stimulus was presented at fixation for 40 ms (see Fig. 1a). A blank screen was presented for 0.5–1.5 s after which a response screen was shown. Observers rated the confidence of their decision (a rating of 1 indicated complete confidence that a face was presented and a rating of 10 indicated complete confidence that a car was presented) by clicking rating buttons on the response window. Participants were instructed only to move the mouse when the response window appeared. Premature mouse clicks were given feedback. No feedback was given about the correct status of their decisions.

#### *Stimuli and display*

The stimulus set consisted of 12 faces (six frontal view, six 45° rotated) and 12 cars (six frontal, six 45° rotated) that were 290 × 290 pixel 8-bit gray scale images. Face images were obtained from the Max Planck Institute for Biological Cybernetics face database (Troje and Bülthoff, 1996). All images were filtered to achieve a common frequency power spectrum (the average of all images). Contrast energy of all images was matched to be 0.3367 deg<sup>2</sup>. Noise was created by filtering independent white Gaussian noise fields (standard deviation of 3.53 cd/m<sup>2</sup>) by the common power spectrum and then added to the original stimuli. Observers sat 125 cm from a linearized display set at a mean luminance of 25 cd/m<sup>2</sup> and maximum of 50 cd/m<sup>2</sup>. Images subtended 4.57° of visual angle and were



**Fig. 1.** (a) Timeline of a trial: a fixation screen with varying presentation time is followed by the test image (stimuli: car or face) presented for 40 ms and followed by a blank screen. Observers are only allowed to make a response after the end of the presentation of the blank screen (500 ms to 1500 ms post-stimuli). (b) Behavioral performance of the twenty individuals and of a multivariate pattern classifier using each individual's neural activity to predict the category of the stimulus (car vs. face). (c) Event Related Potentials (ERP) for face and car trials for the PO8 electrode averaged across observers. (d) Multivariate pattern classifier used to combine individual EEG activity across time and electrodes to make a decision of whether a car or face was presented. The performance of the classifier can then be compared to behavioral performance provided by the ratings.

presented on a 19-inch ViewSonic Color E90F monitor (resolution 1024×768) with a refresh rate of 75 Hz in a darkened room.

#### Electroencephalogram data acquisition and pre-processing

Electroencephalographic (EEG) activity was recorded from 64 Ag/AgCl sintered electrodes mounted in an elastic cap and placed according to the International 10/20 System. The horizontal and vertical electro-oculograms (EOG) were recorded from electrodes placed 1 cm lateral to the external canthi (left and right) and above and below each eye, respectively. The data were sampled at 512 Hz, re-referenced offline to the signal recorded from the central midline electrode (Cz), and then band-pass filtered (0.01–100 Hz). Trials containing ocular artifacts (blinks and eye movements) detected by EOG amplitudes exceeding  $\pm 100$  mV or by visual inspection were excluded from the analysis. The average ERP waveforms in all conditions were computed time-locked to stimulus onset and included a 200 ms pre-stimulus baseline and 500 ms post-stimulus interval.

#### Feature reduction for each individual

The entire EEG data for a trial consisted of 358 time potentials for 63 different electrodes resulting in a total of 22,554 independent inputs. Because of the high dimensionality, for a training set (900 trials of EEG data) 17 electrodes with the highest t-values were pre-selected for a single observer: F1, F5, C5, TP7, P7, P9, PO7, Iz, POz, Afz, F8, FCz, CP6, P8, P10, PO8, and PO4. The same 17 electrodes were then used for all data sets and observers. Analysis of pre-stimulus activity and of simulated data confirmed that no systematic bias in pattern classifier performance resulted from the feature selection.

#### Pattern classifier for separate EEG windows of 20 time-points

For each observer the pattern classifier was trained on windows with 20 time points (at 512 Hz sampling rate each window spanned 39.05 ms) for 17 electrodes. Training was done on 900 trials of EEG data and then tested on the remaining 100 trials. We used a linear pattern classifier (Duda et al., 2000; Eqs. (1)–(3)) applied separately



to non-overlapping time windows with different onsets ( $t_0 = -166, -127, -88, -49, -10, 30, 68, 108, 147, 186, 225, 264, 303, 342, 381, 420$  and  $459$  ms where  $0$  ms is stimulus onset). For each window the training of the classifier was restricted to data from that specific time window.

#### Combining EEG data across windows

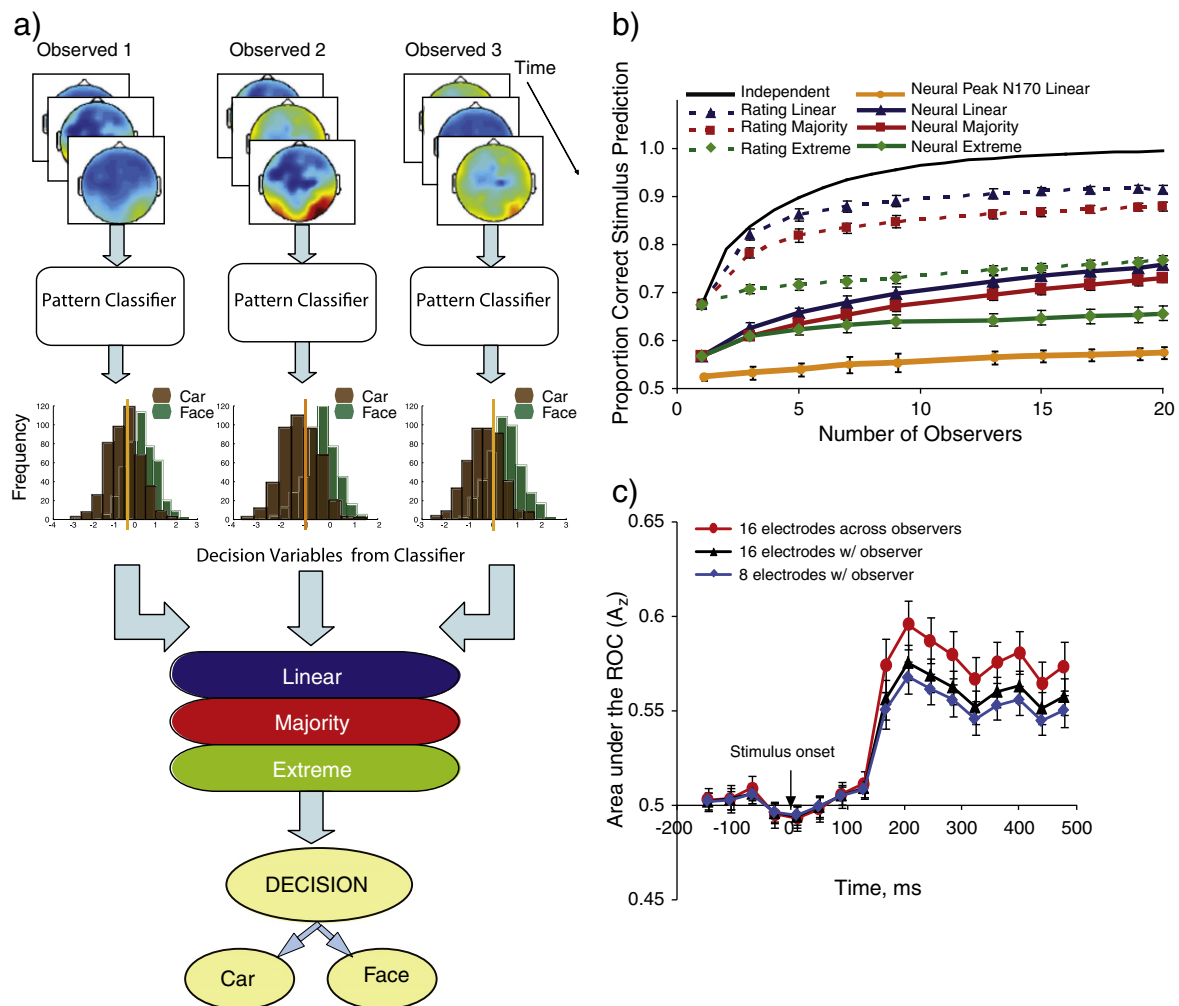
To obtain a final classifier performance across 10 different windows with onsets from  $108$  ms to  $459$  ms, the scalar values of the classifier from the individual windows from the training data set were used as an input to a 2nd pattern classifier that combined the scalar values of the classifiers applied to individual windows. The weights across the 10 windows were used to reduce for each trial the EEG activity for each observer to a single scalar neural decision variable and obtain a performance for the classifier across all window epochs. The main results and dependence with number of observers (Fig. 2b) remained similar when using a pattern classifier or averaging to combine variables across time windows.

#### Peak N170 amplitude and mean amplitude from PO8 electrode

We compared the pattern classifier method with a more traditional method that extracted for each trial the peak and mean amplitude from the time course of EEG close to the N170 component. The peak amplitude and peak latency were measured using an automatic peak detection routine within a window of  $140$ – $210$  ms after onset of the stimulus on the EEG. For each subject and trial, mean amplitude N170 was calculated using a  $\pm 40$  ms window centered on the peak N170 latency.

#### Combining behavioral ratings and neural decision variables across individuals

The behavioral confidence ratings for different observers for the  $i$ th image were combined to make group decisions using: a) optimal linear combination; b) majority rule; c) extreme decision rule. The neural decision variables for the observers for the same  $i$ th image were also combined using the same decision rules (Fig. 2a).



**Fig. 2.** (a) Procedure to reduce each observer's neural activity to a neural decision variable that can be linearly combined across individuals, reduced to a binary decision (car or face) with a neural decision criterion before using the neural majority rule (evaluated for odd number of observers) or used to determine the observer with the most extreme neural response relative to their neural decision criterion (see Materials and methods). (b) Performance discriminating cars from face image trials for various decision rules used to combine information across observers (optimal linear, majority and extreme) as a function of group size (number of observers) for behavioral ratings (dashed lines) and for the neural decision variables extracted from the EEG activity (continuous lines) ranging from  $100$  to  $500$  ms after stimulus onset (see Materials and methods for error bar calculation using bootstrap re-sampling). Lower yellow continuous line is classifier performance for an optimal linear combination of mean amplitude around the peak of single trial N170 EEG. Continuous black line corresponds to performance improvement expected if observers were independent detectors<sup>4</sup> (see Materials and methods); (c) performance for linear classifier discriminating images of cars from faces for different  $39$  ms windows with varying time onsets using varying combinations of electrodes and observers:  $8$  electrodes within one observer,  $16$  electrodes within one observer, two sets of  $8$  electrodes from different observers.

### Optimal linear combination

A single scalar decision variable ( $D_i$ ) was obtained by linearly combining the ratings or neural decision variables for each trial (image) across observers,

$$D_i = \mathbf{w}^t \mathbf{R}_i \quad (1)$$

where  $\mathbf{w}$  is a column vector containing the linear weights applied to the vector  $\mathbf{R}_i$  containing the ratings or neural decision variables of the observers and the superscript  $t$  is transpose. This algorithm is optimal when the underlying distributions are equal variance and Gaussian distributed. The decision outcome was determined by comparing the decision variable ( $D_i$ ) to a scalar decision criterion ( $C_i$ ; the subscript  $i$  denotes linear) that maximized proportion correct:

$$O_i = \begin{cases} 1, & \text{if } D_i < C_i \\ 2, & \text{if } D_i > C_i \end{cases} \quad (2)$$

where  $O_i = 1$  indicates a choice for the face category and  $O_i = 2$  a choice for the car category.

For each random sample of  $N$  observers, the behavioral ratings/neural decision variables for 900 trials were used for the training and performance was obtained in the remaining 100 trials. The cross-validation procedure was repeated 10 times. The best linear weights ( $\mathbf{w}$ ) for the observers were estimated from the training data:

$$\mathbf{w} = \mathbf{K}^{-1} [\langle \mathbf{R}_f \rangle - \langle \mathbf{R}_c \rangle] \quad (3)$$

where  $\langle \mathbf{R}_f \rangle$  is a column vector containing the mean rating/neural decision variable for the face images for the  $N$  observers,  $\langle \mathbf{R}_c \rangle$  is a column vector containing the mean rating/neural decision variable for car images for the  $N$  observers, and  $\mathbf{K}^{-1}$  is the inverse of the covariance matrix which is calculated from the ratings from training images (900 trials):

$$\mathbf{K} = \langle (\mathbf{R}_f - \langle \mathbf{R}_f \rangle)(\mathbf{R}_f - \langle \mathbf{R}_f \rangle)^t \rangle + \langle (\mathbf{R}_c - \langle \mathbf{R}_c \rangle)(\mathbf{R}_c - \langle \mathbf{R}_c \rangle)^t \rangle. \quad (4)$$

In order to stabilize the covariance across all procedures, invert  $\mathbf{K}$  and avoid singular matrices, we regularized the matrix by adding a diagonal matrix with its diagonal elements proportional to the diagonal elements of the covariance matrix  $\mathbf{K}$ :

$$\mathbf{K}_{\text{reg}} = \mathbf{K} + p \text{Diag}(\mathbf{K}) \quad (5)$$

where  $p$  is a proportional constant set to 2 and the function  $\text{Diag}$  zeroes all off-diagonal elements of  $\mathbf{K}$ . Results did not vary greatly with varying values of  $p$  (1, 2, 3 or 4).

The decision criterion ( $C_i$ ) was also estimated from training data (900 trials) and applied to the 100 trials left out.

### Majority decision rule

For each trial the number of individuals choosing one category (e.g., face) can be expressed:

$$M_i = \sum_{j=1}^n \text{step}(R_{j,i} - c_j) \quad (6)$$

where  $R_{j,i}$  is the rating/neural decision variable for the  $i$ th trial/image for the  $j$ th individual ranging from 1 to 10.  $c_j$  is the criterion which for the ratings is the value that subdivides the categories corresponding to face and those to car ( $c_j = C = 5.5$ ). For the neural data, the neural decision criterion ( $c_j$ ) was varied for each  $j$ th individual to find the value that maximizes the choice probability. Once the criterion was determined from the training data a neural majority decision was implemented for each trial of the testing data by first making binary

decisions for each individual based on their neural decision variable and neural decision criterion and then using Eq. (8). Also, *step* is the heavyside step function defined as:

$$\text{step}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (7)$$

The outcome of the majority decision is then given by:

$$O_i = \begin{cases} 1, & \text{if } M_i < (n-1)/2 \\ 2, & \text{if } M_i > (n-1)/2 \end{cases}. \quad (8)$$

The above equations work for most cases considered in this paper with odd number of individuals and do not cover the cases in which there are an even number of individuals and a possibility of a tie between group members choosing the two possible decisions. For the one such case we analyzed (e.g.,  $n = 20$ ), if there was a tie, we simply randomly chose a decision.

### Extreme decision rule

For each trial the group decision was based on the observer whose response (rating or neural decision variable) departed the most from the decision criterion ( $C$ ). For the behavioral data the subdivision between the two categories ( $C = 5.5$ ) was set as the criterion. For the neural data we evaluated for each trial which observer had a neural decision variable that departed the most (in units of standard deviations) from their neural criterion ( $c_j$ ).

$$D_i = \max_{j=1}^N (R_{j,i} - c_j) \quad (9)$$

where the  $\max_{\text{abs}}(x)$  function takes the response  $R_{j,i}$  with the largest absolute value difference from the criterion,  $c$ . The response outcome  $O_i$  is then determined using Eq. (2). Note that Eq. (9) does not handle the possibility for the behavioral ratings in which there is a tie between the most extreme decision choosing car and face (e.g., an observer chooses a rating  $R_j = 1$  and another in the group chooses  $R_j = 10$ ). For these tie cases we compared the number of observers choosing each of the two extreme ratings and chose the category with the highest number of observers. If this was still a tie, we counted the number of observers in the 2nd most extreme rating (e.g.  $R_j = 2$  vs.  $R_j = 9$ ).

### Prediction for performance improvement if observers were independent detectors

For comparison to empirical results we calculated the predicted performance (proportion correct) with number of observers for an optimal combination across independent detectors. The theoretical prediction relies on the assumption that each observer's responses to car and face trials are described by an equal variance Gaussian model in which the index of detectability,  $d'$ , is the difference between the mean of the distributions divided by their common standard deviation. The index of detectability as a function of the number of observers ( $n$ ) for the case of varying indexes of detectability across observers is calculated as (Sorkin & Dai, 1994):

$$d'_n = \sqrt{n(\sigma_{d_i}^2 + \bar{d}^2)} \quad (10)$$

where  $\bar{d}$  is the mean index of detectability across all observers and  $\sigma_{d_i}^2$  is the variance of the observers' indexes of detectability. These theoretical predictions from Eq. (10) assume that a very large number of trials are available to train the classifier and thus represent an upper-bound. To obtain a proportion correct prediction from the predicted  $d'_n$  we simply

assumed an optimally placed criterion ( $d'_n/2$ ). Proportion correct identification is then given by:

$$P_c = \int_{-\infty}^{d'_n/2} g(x) dx \quad (11)$$

Note, that in the theoretical prediction in Fig. 2a there is an additional increase in performance from  $n=1$  to  $n=2$  due to the fact that for  $n>1$ , the decision criterion is optimally placed while individual observers ( $n=1$ ) might use suboptimal criteria.

#### *Effect of correlation across observer responses on the relationship between performance of optimal linear combination and majority decision rules*

We performed a simulation sampling for each trial and observer random variables from two normal distributions (with unit variance). For a face trial the normal distributions were assumed to have a larger mean and the car trials a zero mean. Means for the simulated twenty observers for the face trials were matched to observers' index of detectability from the study. Critically, the sampled random variables were given a pre-determined correlation which for simplicity was assumed to be constant across observers. All procedures to combine responses were identical to those in the paper except that the simulation results were based on 1000 different group configurations. We varied the correlation across observer's simulated decision variables and evaluated the relationship between performance of the optimal linear combination and the majority decision rule.

#### *Cumulative integration of EEG data across time*

To obtain a classifier performance using neural activity from stimulus onset up to  $T$  ms, we first ran separate classifiers across 10 ms temporal windows. We then obtained the performance of integrating the classifier decision variables across windows ranging from 0 to  $T$  ms where the upper limit of the integrating window ( $T$ ) was systematically increased by 10 ms.

#### *Theoretical calculation of classifier performance from cumulative integration of EEG data across time for the scenario of temporally independent samples*

We compared our empirical results to theoretical calculations of performance of the multivariate pattern classifier expected if the noise values in the neural activity (EEG) were statistically independent samples through time. The theoretical calculation assumes that mean difference in classifier decision variables across the two stimuli categories remained constant across time and the noise was temporally statistically independent.

To estimate performance of the classifier that integrates neural decision variables across for this theoretical scenario we used the well-known relationship between the index of detectability ( $d'$ ) and the number of independent samples integrated (Green and Swets, 1989; Eckstein et al., 1996)

$$d'_{N_T} = d'_0 \sqrt{N_T} \quad (12)$$

where  $d'_{N_T}$  is the estimated index of detectability (mean difference between the decision variable to face and car divided by a standard deviation of the decision variable) at time  $T$  after integrating  $N_T$  independent samples of 10 ms of EEG activity and  $d'_0$  is the average index of detectability of the pattern classifier in the first temporal sample which is statistically above chance (70 ms). To estimate the

initial index of detectability,  $d'_0$  we used the proportion correct from our data for each group size and transformed it to an index of detectability using the following relationship (Green and Swets, 1989),

$$d'_0 = 2 \cdot \Phi^{-1}(P_{c_0}) \quad (13)$$

Where  $\Phi^{-1}$  is the inverse normal cumulative function. Eq. (12) was used to obtain  $d'_{N_T}$  from  $d'_0$ . Finally to obtain  $P_{c_{N_T}}$  from  $d'_{N_T}$ , we used Eq. (11).

#### *Controls to ensure no over-fitting of the pattern classifier to the data*

To ensure that none of our procedures (feature selection) led to over-fitting we did not train using any EEG activity from earlier epochs of a trial. In addition, we trained and tested the classifier on pre-stimulus presentation data with the same procedures as with the post-stimulus presentation data. Pre-stimulus classifier performance at chance confirmed no overtraining in our classifier technique.

#### *Procedures to predict observer choices*

To predict observer choices, the pattern classifier weights were calculated from the training data sets using labels for each trial based on the observer's decision (car or face) rather than the presented image. All other procedures were identical to those used to predict the stimulus type.

#### *Evaluation of performance using Area under the ROC*

Performance evaluation for individual observers of both behavioral ratings and neural decision variables was evaluated in the 100 trials left out using Eq. (1) to compute the scalar value for each trial. Performance as measured by the Area under the Receiver Operating Curve ( $A_z$ ) was calculated using a non-parametric method that quantifies for each value of the pattern classifier to the face stimuli (across  $F$  face trials) the probability that it will exceed the responses to all the pattern classifier values to the car stimuli across all  $C$  car trials in the test data set (DeLong et al., 1988):

$$A_z = \frac{1}{FC} \sum_f \sum_c \text{step}(D_f - D_c) + \frac{1}{2} \delta(D_f - D_c) \quad (14)$$

where  $D_f$  is the scalar decision variable for the  $f$ th face stimuli,  $D_c$  is the scalar decision variable for the  $c$ th car,  $F$  and  $C$  are the total number of face and car trials and  $\text{step}$  is the heavyside step function defined in Eq. (14).

The function  $\delta$  is the impulse function defined as

$$\delta(x) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{if } x \neq 0 \end{cases} \quad (15)$$

The first term ( $\text{step}$  function) inside the summation measures the frequency in which a given classifier response to a face stimuli exceeds the response to the car stimuli. The second term with the impulse function handles the instances in which the pattern classifier values for the face and car stimuli are a tie and the classifier is forced to guess (i.e., the frequency of correct decisions for ties is  $\frac{1}{2}$  the frequency of the ties). Use of a parametric binormal model to calculate the Area under ROC resulted in similar results to the non-parametric method. The training and testing procedure above was repeated 10 times by dividing the total of 1000 trials into 10 groups of 100 trials and then leaving one group of 100 trials for the test and training on the remaining 9 groups with a total of 900 trials. The resulting  $A_z$  from the 10 cross-validations were averaged.

### Evaluation of performance using proportion correct

Because the  $A_z$  measure of performance cannot be computed for the majority decision rule, we also quantified performance using a proportion of correct trials which can be computed for all decision rules as:

$$P = \frac{1}{I} \sum_i \delta(O_i - q_i) \quad (16)$$

where  $O_i$  is the outcome for the  $i^{\text{th}}$  trial (1 or 2),  $q_i$  indicates the stimulus type presented (face=1; car=2),  $\delta$  is as defined before (Eq. (15)) and  $I$  is the total number of trials.

### Estimation of error in measurements of pattern classifier performance on individual observer data

Error bars for the pattern classifier performance on the EEG data of individual observers (Fig. 1b) were calculated by calculating the standard error across Area under the ROC for the 10 fold cross validation sets. Performance was compared to chance (0.5) using single sample t-tests with corrections for multiple comparisons (false discovery rate).

### Estimation of error in measurements of performance after combining multiple observer data

We are interested in comparing performance on the same data set across groups of different numbers of observers. To calculate the error due to observer variability and group configuration we used bootstrap methods to calculate the uncertainty associated with the mean performance of a classifier for 400 samples of groups of  $N$  observers. Error bars were standard deviations calculated across bootstrap samples of observers calculated as follows:

- We created one hundred bootstrap lists of 20 observers.
- For each bootstrap list we randomly sampled four hundred groups of  $N$  observers.
- For each random group we combined across observers using the specific decision rule (linear, majority, most extreme).
- We computed the average Area under the ROC or proportion correct for the 400 random samples for each bootstrap sample.
- We computed the standard deviation of the average Area under the ROC/proportion correct curve across bootstrap samples.

One potential problem in implementing the bootstrap technique within the context of combining ratings/neural activity across observers with a linear pattern classifier is that having the same observer twice represented in the covariance matrix will lead to a singularity. However, the regularizer (see Eq. (4)) allows for inversion of the matrix when the same observer is repeated in the covariance.

### Across brain vs. within brain pattern classifier performance

We also evaluated the benefits of integrating information across multiple brains while controlling for the data size used. We divided the electrodes into two groups to compare the effect of combining two groups of electrodes within a brain vs. across two brains. This analysis required an even number of total electrodes so that each group had the same number of electrodes. Thus for this analysis, a subset of sixteen out of the seventeen electrodes was chosen for a subsequent analysis to compare within brain vs. across brain pattern classifier performance. We then compared the improvement in the classifier's performance discriminating car from face trials when combining the neural activity from the two groups of eight electrodes within the same brain vs. the two groups of electrodes across two different brains. The procedure was repeated 100 times by randomly sampling

two brains and dividing the sixteen electrodes into two random groups of eight electrodes.

## Results

### Behavioral and pattern classifier performance of individual observers

Perceptual performance (Area under the Receiver Operating Curve) identifying the object in the images varied widely across observers ranging from 0.581 to 0.893 (Fig. 1b). Fig. 1c shows the EEG data aligned with respect to stimulus presentation and averaged across all trials separately for cars and faces (event related potential, ERP). The results show the typical negative-going deflection in the ERP component differentiating responses to face stimuli relative to other objects (i.e., cars) and peaking around 170 ms after stimulus onset (Bentin et al., 1996; Itier and Taylor, 2004; Rousselet et al., 2008; for more details on the N170 with and without external noise; Figure S1.). To relate the neural activity to task performance we used the EEG activity from each trial and multiple electrodes (see Materials and methods). The classifier was trained using a 10 fold cross-validation procedure to discriminate the neural activity evoked by faces and cars from individual observers (Fig. 1d, right flow diagram) and resulted in a decision variable for each trial that was used to make choices about the image (car vs. face; Fig. 1b). Classifier performance operating on observers' neural activity positively correlated with their behavioral performance ( $r = 0.751$ ;  $p < 0.01$ ).

### Pooling behavioral decisions across observers

To explore the benefits of collective wisdom in our task, we aggregated observers' behavioral ratings across different size groups ( $N = 3, 5, 7, 9, 13, 15, 17, 19$  and 20) using an optimal linear combination (Duda et al., 2000). Because the impact of combining ratings across a group of individuals will likely depend on the particular individuals in the group, we repeated this and all analyses using the ratings from four hundred random group configurations of  $N$  individuals sampled from our list of twenty human observers. Fig. 2b shows the typical improvement in proportion correct with group size but less than that expected if observers were independent detectors (see methods for mathematical details; Green and Swets, 1989). We also investigated a majority rule for which the group decision corresponded to the choice selected by the greater percentage of individuals. This simpler majority decision rule across observers showed substantial benefits of combining behavioral ratings across observers but was lower than the optimal linear (Fig. 2b). Finally, a group decision rule that relied on the groups' most extreme rating (see Materials and methods) results in much smaller improvements with increasing number of individuals.

### Pooling neural decision variables across multiple brains

To compare the benefits of aggregating behavioral ratings across individuals to the combination of the neural activity across their multiple brains, we first used the multivariate pattern classifier to reduce the multi-electrode EEG activity (100–500 ms post stimulus) of each individual and for each trial to a neural decision variable (Fig. 2a). We then combined for each trial the neural decision variables across brains using an optimal linear combination (see Materials and methods). Fig. 2b shows that the performance (proportion correct) after integration of the neural decision variables across multiple brains increases with group size and parallels the benefits of linearly combining the observers' behavioral ratings. However, performance of the pattern classifier based on neural activity is lower than that based on the behavioral ratings suggesting that our method cannot recover from the EEG activity all of the task-based information inherent in the behavioral ratings. To match the



mean behavioral accuracy of a single observer, the neural classifier required integrating EEG activity across seven brains. To evaluate a neural majority and a neural most extreme opinion decision rules, we determined for each individual a neural decision criterion that maximized the probability of predicting their trial-by-trial decisions (see Fig. 2b). The neural decision criterion allowed us to categorize the neural decision variables into car and face decisions for each trial and compute a majority decision (Fig. 2b).<sup>2</sup> Use of the neural majority decision rule required combining the EEG of nine brains to match behavioral performance of the mean single observer.

A neural most extreme opinion decision rule was implemented by making group choices for each trial solely based on the individual whose neural decision variable was most distant from their own neural decision criterion. To implement this, we quantified for each trial and individual the neural decision variable's distance in standard deviation units from the observers' neural decision criterion. We then determined the trial group decision based on the individual with the highest amplitude deviation from the decision criterion. The relationship among the performances of the three neural decision rules is similar to their relationship when applied to the observer ratings (Fig. 2a). In particular, the neural majority rule was close to the linear combination rule and the neural extreme opinion was highly suboptimal.

Finally, we compared the result of integrating neural activity across brains preceded by using the pattern classifiers on individual brains to that of a more conventional methods common in the ERP literature such as using a peak amplitude and mean amplitude around the N170 component. For each observer we extracted for each trial the peak and mean amplitude in a 40 ms window from the PO8 electrode (see Materials and methods). The peak or mean amplitude for each trial from multiple brains was then used as inputs to the optimal linear and majority rules to combine mean/peak amplitudes across observers. Fig. 2b shows that the performance achieved by combining peak amplitudes of the N170 across observers using an optimal linear rule is much lower than that achieved by first extracting from each observer a neural decision variable from applying a pattern classifier to a group of electrodes and all time-points. Results for the N170 mean amplitude and majority decision rule were similar or lower to the N170 peak amplitude and are shown in Supplementary Figure S2.

#### Unique benefits of integrating information across multiple brains

Although the results in the previous section are consistent with the hypothesis that the classifier's benefit in performance with increasing number of brains is associated with a property unique to combining neural activity across different brains, it is also possible that it solely reflects a simple increase on the available data size used for the classifier irrespective of whether the data belongs to a single brain or different brains. To test this alternative we divided the electrodes of two individuals into random groups of eight electrodes. We then compared the improvement in the classifier's (optimal linear) performance discriminating car from face trials when combining the neural activity from the two groups of eight electrodes within the same brain (16 electrodes from one observer) vs. two groups of eight electrodes across two different brains (8 electrodes from two observers). If there is a unique benefit associated with integration of information across two different brains, then the classifier's performance should be higher when combining across the two groups of electrodes from two different brains than when combining within single brains. Because we want the result not to be dependent on the particular two brains and two groups of electrodes selected, we repeated the procedure 100 times by randomly sampling two brains

and dividing the sixteen electrodes into two random groups of eight electrodes. The same procedure was implemented for various combinations of number of electrodes and number of brains while keeping the data size constant (total of 16 electrodes). Fig. 2c shows that performance ( $A_z$ ) of the pattern classifier increases as it integrates neural activity across larger number of different brains (16 electrodes from 2 brains vs. 16 electrodes from a single brain;  $p < 0.05$  with a correction for multiple comparisons). The result holds for all time windows following 147 ms after stimulus presentation. For comparison Fig. 2c also shows the performance improvement resulting from using a larger number of electrodes from a single observer (16 electrodes from a single brain vs. 8 electrodes from a single brain).

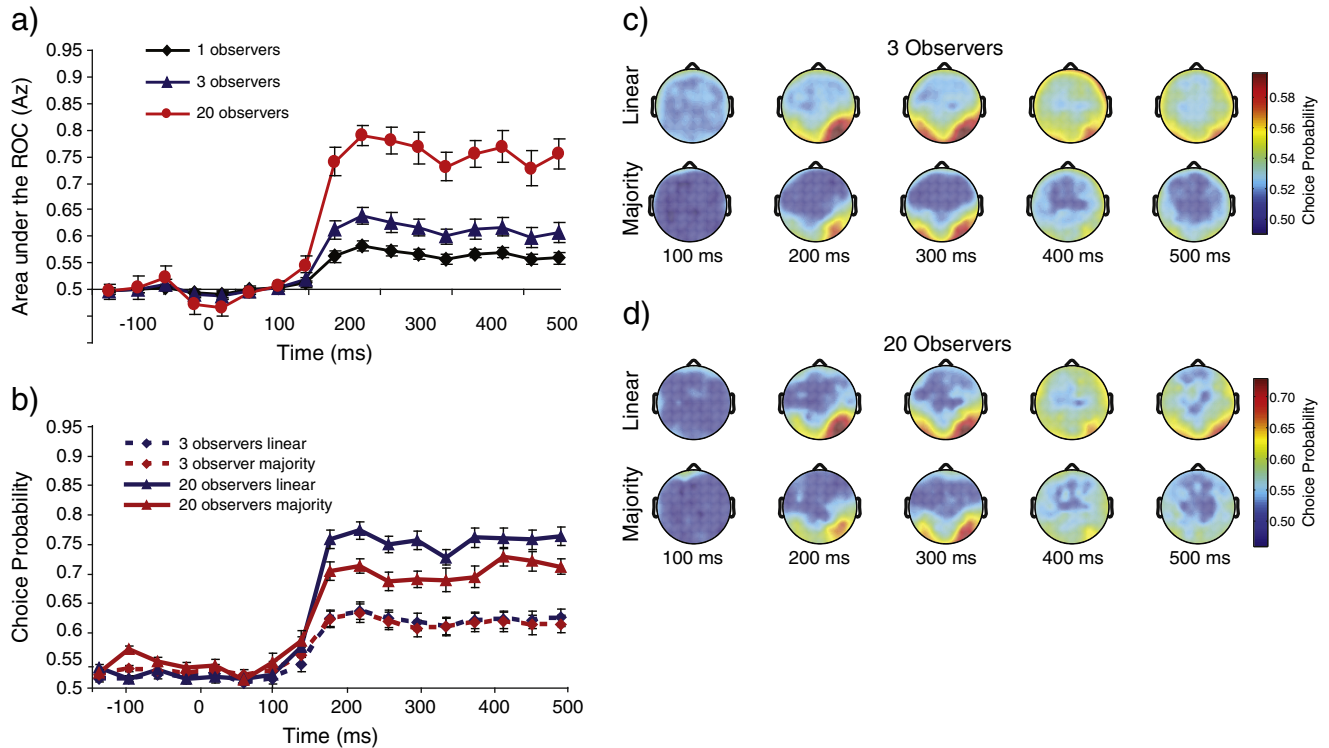
#### Performance benefits from integrating neural decision variables in early sensory vs. late decision stages

To assess whether integration across brains of neural activity associated with the late decision component leads to greater performance benefits than from combination of early sensory activity, we measured accuracy ( $A_z$ ) after optimal linear pooling across brains of neural decision variables for different temporal windows of 39 ms. Fig. 3a shows accuracy improvements with number of brains ( $N = 1, 3$ , and  $20$ ) as a function of time. With the EEG activity prior to the presentation of the stimulus accuracy remains at chance ( $0.5$ ), but increases significantly starting with a temporal window of EEG activity centered around 167 ms after stimulus onset (Fig. 3a). This onset corresponds to the typical time window of the face-selective N170 ERP component that has been tied to the perceptual coding of faces (Bentin et al., 1996; Itier and Taylor, 2004; Rousselet et al., 2008). The high accuracies persist all through 300–500 ms after stimulus onset, a temporal epoch which has been related to decision components (Philiastides and Sajda, 2006a). Critically, the performance differences across varying number of brains integrated ( $N = 1, 3$ , and  $20$ ) remain constant across the early sensory and late decision-related EEG components ( $p > 0.05$ ).

#### Choice probabilities of groups for early sensory vs. late decision neural activity

If behavioral decisions are governed by signals and noise at the late decision stage we should expect that collective choice probabilities associated with neural activity in the late component will be higher than that related to the early sensory stages. Fig. 3b shows the choice probability analysis for the various temporal epochs for the optimal linear and majority pooling rules. The pattern of results is consistent with the analysis quantifying the accuracy of the EEG predicting the stimulus (Fig. 3a). Choice probabilities increased significantly above chance starting with a temporal window of EEG activity centered around 167 ms after stimulus onset. Importantly, the results show that the choice probabilities predicting the decisions reached by aggregating observer opinions (optimal linear or majority rule) are no better for the neural activity associated with the late decision component than for that associated with the early sensory stage. Furthermore, the topographic layout of choice probabilities across electrodes and time provides additional data. Although the spatial localization of the neural activity is limited by the spatial resolution of EEG and the inverse problem, the scalp topographies (Figs. 3c and d) show highly significant choice probabilities over lateral occipital cortex starting around the onset of the N170 ERP component with more spatially distributed significant choice probabilities at later times. As a final comparison, choice probabilities resulting from using only the peak or mean amplitudes of the N170 of the PO8 electrode as inputs to the pattern classifier did not reach values higher than  $0.54 \pm 0.03$ .

<sup>2</sup> For the case of  $N = 20$  the ties ranged between 1 and 2% of the trials. Ties only affected this one data point ( $N = 20$ ) and do not affect our main results (see Materials and methods for handling of ties).



**Fig. 3.** a) Performance ( $A_z$ ) predicting the stimulus identity (car vs. face) as a function of number of brains for neural linear decision rule for different 39 ms windows with varying time onsets. Figure shows conditions for one brain (averaged across all twenty brains), and groups of three brains and all twenty brains and averaged across 400 samples of groups of three brains; b) choice probability (see [Materials and methods](#)) of the neural linear and neural majority decision using the human EEG data for different 39 ms windows with varying time onsets; c) topographic maps of choice probabilities for different time points for linear and majority combination rules and averaged across 400 random configurations of three observers; d) same as previous but for groups of twenty observers.

#### Spatiotemporal distribution of correlations between group behavioral decision confidence/voting outcome and neural activity

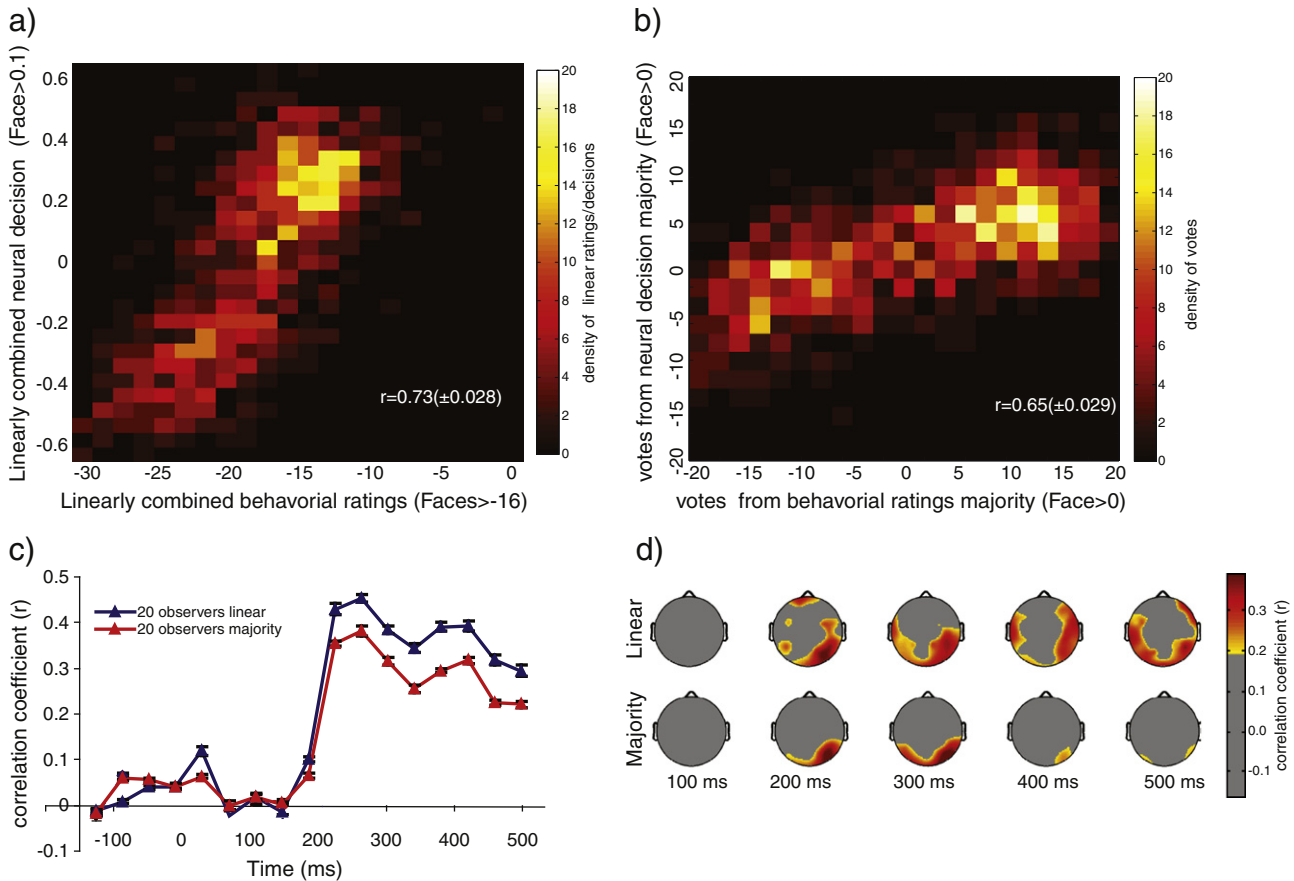
If the measured neural decision variable reflects group evidence for or against the presence of a car or face, we expect a significant positive correlation between confidence ratings of the weighted average of the observers' opinions and the multi-brain integrated neural decision variable. [Fig. 4a](#) shows a scatter plot of the data and a significant correlation ( $r = 0.7$ ;  $p < 0.0001$ ) between the optimal linear combination of behavioral confidence ratings and that of the integration of the neural decision variables of all twenty brains. Similarly, although the majority rule does not provide an explicit confidence rating for a decision one can interpret the voting outcome as an expression of confidence (e.g., 90% agreement in a decision as more confident than a decision with close to an even split across voters). [Fig. 4b](#) shows a positive correlation between the voting outcome obtained from the behavioral opinions and the votes obtained using the neural decision variables from all twenty brains ( $r = 0.64$ ;  $p < 0.0001$ ). For these analyses the neural decision variable is integrated across the entire EEG time epochs.

Critical to one set of hypotheses tested in the present paper, if the late decision component governs the formation of choices and determines the decision confidence we might expect that decision related neural activity (300 ms post stimulus onset) will be more highly correlated with the pooled behavioral confidence than that associated with sensory activity. However, the time course of the correlation between the pooled behavioral confidence rating and the pooled neural decision variables shows no higher correlations for the late decision component than the early sensory activity. In fact, the correlation peaks at 250 ms after stimulus onset ([Fig. 4c](#)). Finally, the scalp topography of the correlations between behavioral pooled confidence ratings and neural decision variables shows that statisti-

cally significant high correlations are localized at lateral occipital electrodes starting at 200 ms after stimulus onset.

#### Incremental temporal integration of EEG activity and faster potential decisions when combining neural decision variables across brains

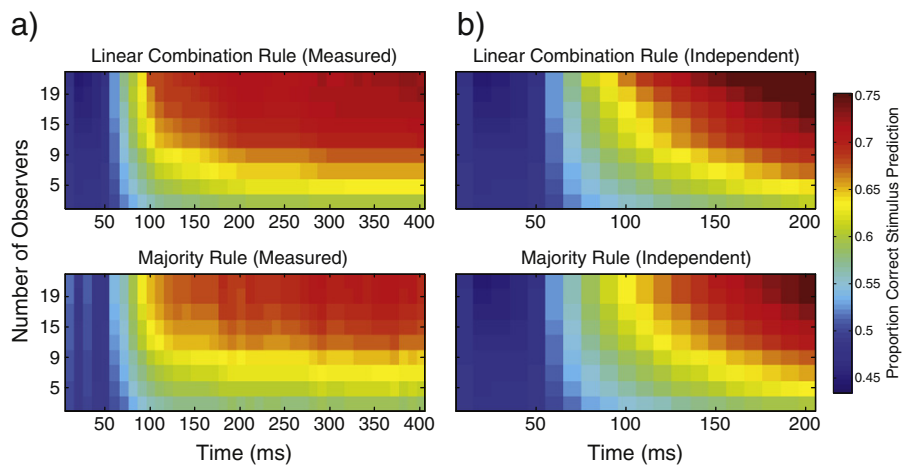
We evaluated the relationship between integration of neural activity across multiple brains and the time-course of neural activity discriminating the two perceptual stimuli. We first grouped for each observer EEG time samples into bins of 10 ms ([Fig. 5a](#)) and trained a classifier to maximally discriminate between the car and face stimuli. The neural decision variables for each temporal bin were summed within an integration window that was incrementally extended from 0 ms to 400 ms. The integrated neural decision variables were then combined across multiple brains using an optimal linear combination or using a majority rule. [Fig. 5a](#) plots accuracy of temporally integrated pattern classifiers' neural decision variables as a function of aggregation across varying number of brains. The results show that combining information across a larger number of brains, either through a linear combination or majority rule, can potentially allow the collective to make faster decisions to achieve a specific accuracy. The measurements provide estimates of relative benefits of combining information across brains vs. those of decision time delays for an individual to temporally integrate more neural information and achieve a higher behavioral accuracy. For example, the measured neural decision variable of a single brain reaches an average task accuracy of 0.56 at 220 ms. On average, the combined neural decision variables of five brains reach that accuracy in less than half the time: 80 ms. These time savings can be compared to what might be theoretically expected if the noise in the EEG activity were temporally independent samples allowing for larger performance improvements through temporal integration of single brain activity ([Fig. 5b](#)). Theoretical predictions were obtained assuming independent



**Fig. 4.** a) Two-variable histogram of linearly aggregated behavioral opinions (ratings) of 20 observers (x-axis) vs. linear combination of neural decision variables extracted from each brain using a multivariate pattern classifier; b) two-variable histogram of voting results based on observers' behavioral ratings vs. voting results based on brain's neural decision variables; c) correlation between pooled behavioral confidence ratings and neural decision variable using the human EEG data for different 39 ms windows with varying time onsets; d) topographic maps of correlation coefficients for different time points and electrode locations for linear and majority combination rules for groups of twenty observers (gray areas correspond to correlation coefficients which were not statistically significant;  $p > 0.05$  with false discovery rate correction for multiple comparisons).

zero mean noise and a fixed mean difference in EEG across car and face stimuli starting at 70 ms. The mean difference in EEG across stimulus types at the initial time-point was estimated from the first temporal window in the data at which the average single brain pattern classifier performance was statistically above chance (see [Materials and methods](#)

for more details). Based on the theoretical calculations, the temporally independent noise samples scenario would predict that on average a single brain would reach an accuracy of 0.56 at 130 ms rather than 200 ms and thus would predict a smaller time-savings from collective integration ([Fig. 5b](#)).



**Fig. 5.** (a) Left: performance discriminating images of cars from faces from measured neural activity integrated up to t ms (horizontal axis; see [Materials and methods](#)) and combined across groups of N brains (vertical axis) for two different neural decision rules: linear integration (top) and majority (bottom); (b) right: theoretical predictions assuming temporal independent samples of neural activity for performance discriminating images of cars from faces from neural activity integrated up to t ms (horizontal axis; see [Materials and methods](#)) and combined across groups of N brains (vertical axis) for two different neural decision rules: linear integration (top) and majority (bottom).

### Using neural activity to rank order groups of individuals based on the accuracy of their decisions

Can brain activity of individuals be used to reliably identify which of two unique groups would reach a more accurate decision through a simple yet prevalent cooperative decision strategy such as the majority vote? Fig. 6a shows the procedure of using collective neural decision variables to rank order random groups of  $N$  observers based on the performance achieved by aggregating their behavioral opinions. Fig. 6b shows that using the groups' neural majority predictions we can reliably predict with an accuracy of up to 76% ( $n=3$ ) which of two groups would perform the best when aggregating their opinions using a voting and majority scheme or a more complex linear combination. Similar results can be obtained for different group sizes. The decreasing accuracy in assessing which group would best perform the task as group size increases is consistent with the concept that differences in

performance across groups diminish with increasing size due to a regression towards the mean effect. For comparison, ranking of groups based on combination of single trial N170 mean or peak amplitudes resulted in much lower accuracies with maximum values of  $0.543 \pm 0.029$  for the N170 mean amplitude and  $0.542 \pm 0.024$  for the N170 peak amplitude.

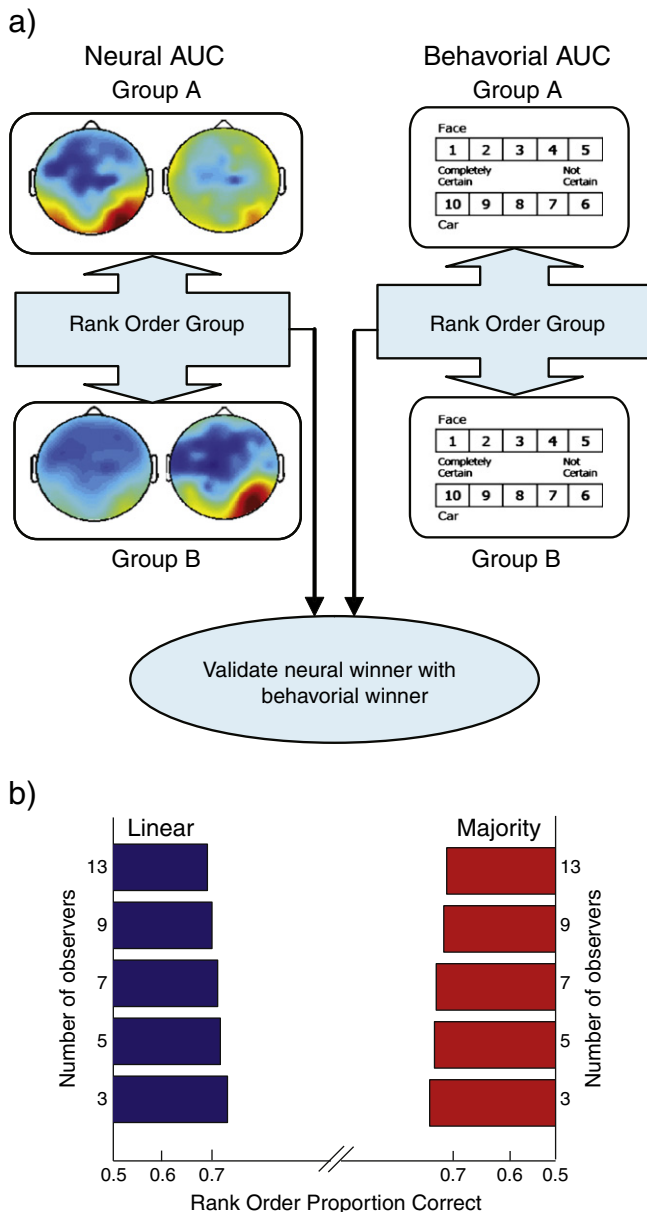
### Discussion

#### Multi-brain computing and improved neural decoding of the state of the world

Over the last decade there has been a dramatic increase in the application of machine learning algorithms using multi-neuron electrophysiology (Quiroga and Panzeri, 2009), functional magnetic resonance imaging (Kay et al., 2008) and EEG (Philiastides and Sajda, 2006a) to read out information from brain activity about the state of the world. Here, we extend these methods to demonstrate that there is increased information about visual stimuli and an impending decision inherent in multiple brains. We show that through multi-brain computing a neural based classifier can reach a performance level at discriminating visual stimuli at an earlier time than using the EEG measurements from a single brain. For a given time-interval integrating neural decision variables across brains leads to a much higher performance identifying the visual stimuli than that achieved with a single brain. An important aspect is that the benefits are not just driven by data size. A control condition showed that, when controlling for data size (total number of electrodes), combining neural activity across different brains brings about unique benefits to the pattern classifier performance. Importantly, our results show that effective multi-brain computing relies in extracting individual neural decision variables using a 1st stage multivariate pattern classifier applied to each individual brain. More traditional schemes such as extracting a single trial peak or mean amplitude around a component for a single electrode leads to much poorer results.

#### Benefits of collective wisdom for perceptual decisions: sensory coding vs. late decision

Improvement in the accuracy and time-efficiency of decisions through collective collaboration is well documented for a variety of animals ranging from bees to humans. For the case of perceptual decisions, the benefits of aggregating individual opinions may come from many sources. First, although all individuals within a species share common sensory anatomy and neural architecture, their different experiences during their lifespan might result in neural mechanisms that use unique subsets of the perceptual information present in the visual environment. Indeed, cell recordings in monkeys have shown how learning modifies the neural activity in visual sensory areas (Gilbert et al., 2001; Li et al., 2004b). Psychophysical studies suggest that learning alters the image features on which humans rely when making perceptual decisions (Li et al., 2004a; Gold et al., 2004) and also that such features vary across human observers (Abbey and Eckstein, 2006). Thus, the sensory mechanisms of different individuals might encode distinct information about a specific environmental event. Integration of neural activity and perceptual judgments across observers might combine the various sources of information represented in different brains leading to better use of information and performance improvement. The inter-observer variability in the coding of stimulus relevant information could occur at the sensory stage but could also possibly occur at a later decision stage. The dorsolateral prefrontal cortex (DLPFC) has been identified as a locus of a domain general mechanism for perceptual decisions that correlates with behavioral performance (Heekeren et al., 2006; Heekeren et al., 2008). For example, there might be less inter-observer variability in the information coded in the sensory



**Fig. 6.** a) Procedure to evaluate the ability to use neural activity to predict which of two mutually exclusive groups of different  $N$  observers would result in more accurate decisions if their behavioral ratings were combined using a majority or a linear decision rule; b) accuracy rank ordering aggregated behavioral opinions of two different groups of  $N$  observers using their neural activity (collective neural decision variable). Figure shows results for linear and majority pooling rules.



stages but the later decision stages (e.g., DLPFC) might sample subsets of the sensory information and lead to greater differences across brains in the coded information. Second, brains are limited by inherent noise at different sensory and decision stages of the transduction of perceptual information that perturbs the neural signals utilized for the decisions (Tolhurst et al., 1983). Thus, aggregation of perceptual judgments across observers might reduce the effects of neural noise resulting in performance benefits. A dominant noise source at the decision stage might also predict greater benefits of collective integration for the late decision component. This is especially the case for perceptual tasks with external image variability (luminance noise) and for which all observers view the same samples of external luminance noise. For such cases, the external noise imposes a correlation across observers' responses diminishing the benefits of integrating information across observers. A large late decision noise source, independent across observers, would further decorrelate the observers' decision variables increasing the gain of integration at decision stages relative to early sensory. Unknown is which of these stages contributes the most towards the benefits of collective perceptual decisions. Here, we compared the performance benefits from integration of neural activity in late decision vs. early sensory stages by analyzing temporal epochs and electrodes of EEG activity. If a decision making mechanism integrates subsets of sensory information in different and unique ways across individuals giving rise to larger variability at the decision stage or if the dominant noise was at the decision stage, then we would expect that the benefits of integration of neural activity to be greater at the decision stage than at the early sensory stage. Yet, our results do not support this hypothesis. Instead, we show that there are no additional benefits in the decision stage beyond what we measure in early sensory activity. The results are consistent with the hypothesis that the benefits of collective perceptual decisions might arise due to variability in the coding of sensory information possibly due to both sensory noise and differences in the subset of visual features utilized by different individuals. It is difficult to pin-point from the present results which visual features might be used by different observers but two possibilities documented in the literature are variability across observers in the use of shape-based features (Gold et al., 2000) contributing to the face/car discrimination and variability in the ability to optimally weight different spatial frequencies. For our images with filtered noise the optimal strategy involves weighting the spatial frequencies in the stimuli inversely to the noise power spectrum resulting in down-weighting of lower frequencies relative to high spatial frequencies. Ability to adjust the weighting of spatial frequencies to non-white noise power spectra has been shown to vary across human observers (Abbey and Eckstein, 2007).

The timing and electrode locus are consistent with the notion that the benefits of aggregating observer opinions for our perceptual task (faces vs. car) possibly arise in regions of the anterior fusiform gyrus and lateral occipital cortex (LOC; Kanwisher and Yovel, 2006). The results however are not inconsistent with the literature showing that the DLPFC is a general purpose decision making mechanism which correlates with behavioral performance and which computes a decision variable using inputs from sensory areas (Heekeren et al., 2008; Heekeren et al., 2006; Philiastides and Heekeren, 2009). Our results might be well-explained if the multivariate pattern classifiers are combining early sensory information in a similar manner to the computations executed in the decision systems. What the present results suggest is that there might not be any different stimulus related information stemming from the integration of activity across sensory areas in decision related areas nor is there late dominant decision noise that governs behavioral choices and decision confidence. Our analysis showing that choice probabilities and correlations with confidence ratings and voting outcomes are as high for early sensory activity as with late decision components is consistent with this interpretation. One assumption in our interpretation is that task-

irrelevant noise does not entirely dominate task-related noise sources in the EEG measurements. In a scenario in which task-irrelevant measurement noise dominated the data then it would diminish our ability to detect any additional performance improvements across early and late processing stages as indexed by EEG. However, an indication that the neural decision variable is not highly dominated by measurement task-irrelevant noise is that the neural decision variables are predictive of behavioral choices. In addition, we found a slight advantage in the correlation with decision confidence of the aggregated group opinions when using earlier EEG activity rather than the later EEG activity, which presumably should not be observed if our measurements were entirely dominated by task-irrelevant noise.

#### *Relationship between collective neural activity and behavioral decision confidence*

The early sensory neural activity is not only related to the binary choices but also correlates with the voting outcome resulting from aggregating observer ratings using a majority vote and the decision confidence after linearly pooling the observers' opinions.<sup>3</sup> A recent study has shown how neurons in the parietal cortex, an area known to be involved in decision making, represent the confidence associated with a perceptual decision (Kiani and Shadlen, 2009). Our results suggest that at least in the human brain, confidence of perceptual decisions and the degree of agreement in a majority vote is not only represented in late decision stages but also in early sensory areas. It is possible that this correlation with decision confidence in early sensory areas might be stronger when the perceptual performance is limited by the sensory data such as the noise in our images. However, even in the presence of external noise, studies have shown that perceptual performance is limited by an equally strong noise internal to the brain (Burgess and Colborne, 1988). Thus, our results suggest that an important component of the internal noise components must be present in the early EEG components associated with sensory processing.

#### *Time-savings by integrating neural activity across brains*

Our analyses also estimated the potential time-savings stemming from integrating neural activity across multiple brains. Our results reveal larger time-savings than we might theoretically expect when integrating information across observers, which stems from the higher correlations of neural activity across time within an observer than across observers. If task relevant neural information was temporally independent, observers could achieve the same performance, as potentially achieved by a group, by shortly delaying their decision and further temporally integrating neural activity. However, the temporal correlations of neural activity require longer time delays to match performance of the group and thus highlight the benefits of collective integration (see Fig. 5a vs. 5b). This interpretation relies on the assumption that the EEG measurement is not introducing the temporal correlations and thus that the measured temporal correlations reflect neural-based correlations.

Speeding decisions by collective pooling requires the animals to have a time-efficient method to behaviorally access the information and integrate it across group members. Indeed, many species including honeybees choosing a new nest site, navigating birds and fish use collective cognition resulting from fast and simple local interactions among group members without complex resource demanding collective behaviors (Couzin, 2009; Sumpter et al., 2008).

<sup>3</sup> Note that using peak or mean amplitude of single trial EEG around the N170 component of a single electrode reached very low choice probabilities.

### Comparing group decision rules applied to neural decision variables and behavioral opinions

We also evaluated a variety of decision rules for integrating neural activity across human brains including an optimal linear, a majority decision, and extreme opinion neural decision rule. Our results show that arguably the most common decision rule in groups of all types of species, the majority rule, benefits nearly as much as an optimal linear rule from pooling task-relevant neural activity across brains. On the other hand a neural decision rule that is based on the brain with most extreme neural decision variable leads to highly suboptimal decision accuracy. The general similarity in the relationship among the behavioral decision rules and their neural analogs might seem inevitable but this might not be necessarily the case. The relationship between the decision rules depends on the correlation among the observers' opinions, among neural decision variables and the response distribution properties. In particular, we conducted simulations creating  $N$  simulated observer decision variables per trial (Gaussian random variables with equal  $d'$  across observers) with varying inter-observer correlation and reached group decisions using the optimal linear and majority decisions (see [Materials and methods](#)). The results confirm that the relationship between the performances of the two decision rules depends on correlation across observers ([Fig. 7](#)). As the correlation increases, the majority rule becomes increasingly suboptimal relative to the optimal linear combination ([Fig. 7](#)). In this context, our results suggest that the correlation structure and distribution properties are at least coarsely preserved across behavioral opinions and neural decision variables. Moreover, despite its computational simplicity, the majority rule can achieve surprisingly high levels of performance under uncertainty ([Quiroga and Panzeri, 2009](#)) and is well-adapted to the variability in inter-observer coding of task-relevant neural activity as assessed through EEG. Others have argued that the majority rule is a robust, simple, efficient social decision heuristic ([Kameda et al., 2011](#)) with low computational cost.

### Practical considerations, applications, and limitations of current study

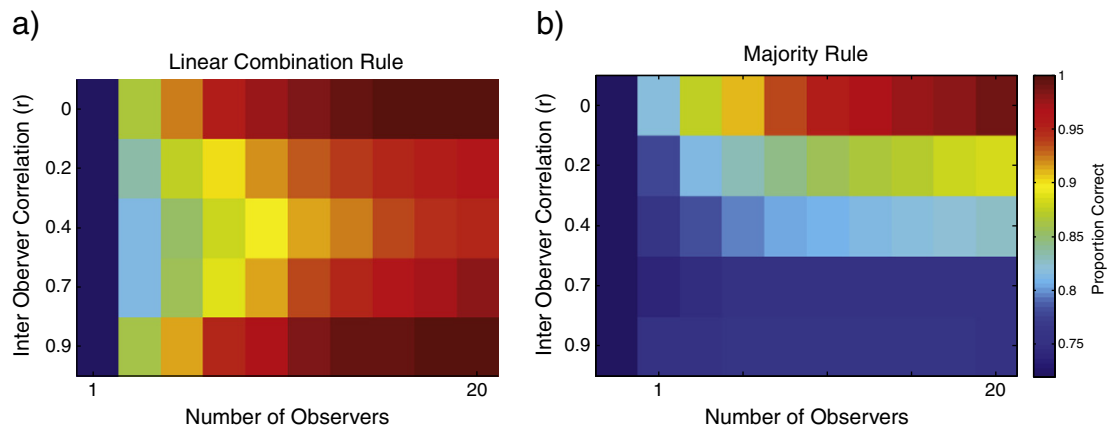
Through multi-brain computing one might be able to develop a neural-based classifier with performance that matches that of the behavior of a single observer. Our results show that the neural classifier required seven brains to achieve performance matched to the mean behavior of a single observer. This result was based on all of our twenty observers, some of which showed poor accuracies for the individual neural classifiers. It is likely that some observers showing consistent superior pattern classifier performance might be better candidates for the EEG technology, and thus a neural-based classifier

based on a select group might require fewer than seven brains to match behavioral performance. The parallel extraction of information from multiple brains allows significant time savings when compared to single-observer EEG which might make the technology particularly applicable for time-sensitive scenarios.

Our results also show that a majority decision rule based on the neural activity of multiple brains can predict the majority decision based on observers' behavioral responses, albeit not perfectly. While encouraging, it is hard to envision scenarios for which the neural voting would replace standard behavioral voting practices. For most important group decisions ranging from judicial to medical decisions there would be no rational reason to replace the behavioral decisions for neural decision, especially, because we presently cannot predict behavioral responses from neural activity without error. Thus, a neural voting approach would perhaps be most appropriate for time-pressured group decisions and scenarios in which humans are engaged in some other motor activity. Neural emergency signals computed in real-time from multiple brains might represent a robust method to detect life-critical situations and even implement automated shutting-off of hazardous machinery. For example, an emergency situation signal computed via neural group decision from the brains of multiple car drivers on a coincident geographic location on a freeway could be one such scenario. A group neural emergency signal could accelerate the deployment of emergency services to the location as well as serve as an input to trigger a response from a vehicle/s (e.g., brake).

Other applications of multi-brain computing include higher performance for cortically coupled computer vision systems ([Sajda et al., 2010; Touryan et al., 2011](#)) and assessments of collective cognitive and emotional states to continuous dynamic stimuli and/or environments. The technology would be limited by the potentially extractable neural correlates of internal cognitive variables through EEG; yet the multi-brain computing framework is potentially applicable to other better measures of neural activity that might be developed in the future.

Finally, the present paper did not compare a large variety of machine learning algorithms and dimensionality reduction methods. Here, we chose a simple set of linear classifiers due to their simplicity and computational speed. Use of more sophisticated algorithms to classify based on iterative optimizations (e.g., support vector machine) or dimensionality reduction techniques (e.g., Independent Component Analysis, ICA, [Jung et al., 2001](#); Class-wise Principal Component Analysis, CPCA, [Das et al., 2007; Das et al., 2010](#)) made some of our sample intensive calculations hard to accomplish in a timely manner. However, future research should concentrate on assessing various algorithms and dimensionality reduction techniques in analyzing the large volume of multi-brain data sets. Arguably, the fact that we obtained the results even with simple



**Fig. 7.** Simulation evaluating group performance as a function of group size with increasing inter-observer correlation (assumed to be equal across observers for simplification) for the optimal linear decision rule (a) and the majority decision rule (b).

linear classifiers is encouraging indicating that improved results might be achieved with more sophisticated algorithms.

## Acknowledgments

This work was first presented at the 2008 Society for Neuroscience Meeting, Washington, D.C. (Eckstein et al., 2008). The project was supported by Army grant W911NF-09-D-0001.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at [doi:10.1016/j.neuroimage.2011.07.009](https://doi.org/10.1016/j.neuroimage.2011.07.009).

## References

- Abbey, C.K., Eckstein, M.P., 2006. Classification images for detection, contrast discrimination and identification tasks with a common ideal observer. *Journal of Vision* 6, 335–355.
- Abbey, C.K., Eckstein, M.P., 2007. Classification images for simple detection and discrimination tasks in correlated noise. *Journal of the Optical Society of America. A* 24, 110–124.
- Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience* 8, 551–565.
- Burgess, A.E., Colborne, B., 1988. Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America. A* 5, 617–627.
- Conradt, L., List, C., 2009. Group decisions in humans and animals: a survey. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364, 719–742.
- Conradt, L., Roper, T.J., 2009. Conflicts of interest and the evolution of decision sharing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364, 807–819.
- Couzin, I.D., 2009. Collective cognition in animal groups. *Trends Cogn Science (Regul. Ed.)* 13, 36–43.
- Das, K., Osechinskiy, S., Nenadic, Z., 2007. A classwise PCA-based recognition of neural data for braincomputer interfaces. *Proc. of the 29th Ann. Int. Conf. of the IEEE Eng: Med. and Biol. Soc.*, pp. 6519–6522.
- Das, K., Giesbrecht, B., Eckstein, M.P., 2010. Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers. *NeuroImage* 51 (4), 1425–1437.
- DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–845.
- Duda, R.O., Hart, P.E., Stork, D.G., 2000. *Pattern Classification*, 2nd ed. Wiley-Interscience.
- Eckstein, M.P., Whiting, J.S., Thomas, J.P., 1996. Role of knowledge in human visual temporal integration in spatiotemporal noise. *Journal of the Optical Society of America. A* 13, 1960–1968.
- Eckstein, M.P., Giesbrecht, B., Pham, B.T., Peterson, M.F., Abbey, C.K., Sy, J., 2008. Neural basis of the benefits of group decisions. Program No. 519.2, 2008. Society for Neuroscience, Washington, DC.
- Galton, F., 1907. Vox populi. *Nature* 75, 450–451.
- Gilbert, C.D., Sigman, M., Crist, R.E., 2001. The neural basis of perceptual learning. *Neuron* 31, 681–697.
- Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. *Annual Review of Neuroscience* 30, 535–574.
- Gold, J.M., Murray, R.F., Bennett, P., Sekuler, A.B., 2000. Deriving behavioural receptive fields for visually completed contours. *Current Biology* 10, 663–666.
- Gold, J.M., Sekuler, A.B., Bennett, P.J., 2004. Characterizing perceptual learning with external noise. *Cognitive Science* 28, 167–207.
- Green, D.M., Swets, J.A., 1989. *Signal Detection Theory and Psychophysics*. Peninsula Pub.
- Hallou, J., Sempo, G., Caprari, G., Rivault, C., Asadpour, M., Tache, F., Said, I., Durier, V., Canonge, S., Ame, J.M., et al., 2007. Social integration of robots into groups of cockroaches to control self-organized choices. *Science* 318, 1155–1158.
- Hastie, R., Kameda, T., 2005. The robust beauty of majority rules in group decisions. *Psychological Review* 112, 494–508.
- Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience* 8, 686–691.
- Heekeren, H.R., Marrett, S., Bandettini, P.A., Ungerleider, L.G., 2004. A general mechanism for perceptual decision-making in the human brain. *Nature* 431 (7010), 859–861.
- Heekeren, H.R., Marrett, S., Ruff, D.A., Bandettini, P.A., Ungerleider, L.G., 2006. Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proceedings of the National Academy of Sciences* 103, 10023–10028.
- Heekeren, H.R., Marrett, S., Ungerleider, L., 2008. The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience* 9, 467–479.
- Itier, R.J., Taylor, M.J., 2004. N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex* 14, 132–142.
- Jung, T.P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.W., Sejnowski, T.J., 2001. Imaging brain dynamics using Independent Component Analysis. *IEEE Proceedings* 88 (7), 1107–1122.
- Kameda, T., Tsukasaki, T., Hastie, R., Berg, N., 2011. Democracy under uncertainty: the wisdom of crowds and the free-rider problem in group decision making. *Psychological Reviews* 118, 76.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8, 679–685.
- Kanwisher, N., Yovel, G., 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 361, 2109–2128.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452, 352–355.
- Kiani, R., Shadlen, M.N., 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759.
- Laughlin, P.R., Bonner, B.L., Miner, A.G., 2002. Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes* 88, 605–620.
- Li, R.W., Levi, D.M., Klein, S.A., 2004a. Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nature Neuroscience* 7, 178–183.
- Li, W., Piëch, V., Gilbert, C.D., 2004b. Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience* 7, 651–657.
- Luo, R.C., Kay, M.G., 1989. Multisensor integration and fusion in intelligent systems. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 901–931.
- Parasuraman, R., 2003. *Neuroergonomics: research and practice*. Theoretical Issues in Ergonomics Science 4, 5–20.
- Neuroergonomics. In: Parasuraman, R., Rizzo, M. (Eds.), *The Brain at Work*. Oxford University Press, 2007.
- Parker, A.J., Newsome, W.T., 1998. Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience* 21, 227–277.
- Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. *NeuroImage* 28, 326–341.
- Philastides, M.G., Heekeren, H.R., 2009. Spatiotemporal characteristics of perceptual decision making in the human brain. In: Dreher, J.-C., Tremblay, L. (Eds.), *Handbook of Reward and Decision Making*. Elsevier, 2009.
- Philastides, M.G., Sajda, P., 2006a. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *Journal of Neuroscience* 26, 8965–8975.
- Philastides, M.G., Sajda, P., 2006b. Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex* 16, 509–518.
- Quiroga, R.Q., Panzeri, S., 2009. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience* 10, 173–185.
- Rousselet, G.A., Husk, J.S., Bennett, P.J., Sekuler, A.B., 2008. Time course and robustness of ERP object and face differences. *Journal of Vision* 8.
- Sajda, P., Parra, L.C., Christoforou, C., Hanna, B., Bahlmann, C., Wang, J., Pohlmeier, E., Dmochowski, J., Chang, S.-F., 2010. A blink of an eye and a switch of a transistor: cortically-coupled computer vision: *Proceedings of the IEEE*, 98, p. 3.
- Seeley, T.D., Buhrman, S.C., 1999. Group decision making in swarms of honey bees. *Behavioral Ecology and Sociobiology* 45, 19–31.
- Shah, S., Aggarwal, J., Eledath, J., 1997. Multisensor integration for scene classification: an experiment in human form detection. *ICIP*, p. 199.
- Simons, A.M., 2004. Many wrongs: the advantage of group navigation. *Trends in Ecology & Evolution* 19, 453–455.
- Sorkin, R.D., Dai, H., 1994. Signal detection analysis of the ideal group. *Organizational Behavior and Human Decision Processes* 60, 1–13.
- Sorkin, R.D., West, R., Robinson, D.E., 1998. Group performance depends on the majority rule. *Psychological Science* 9, 456–463.
- Sorkin, R.D., Hays, C.J., West, R., 2001. Signal-detection analysis of group decision making. *Psychological Review* 108, 183–203.
- Sumpter, D.J., Krause, J., James, R., Couzin, I.D., Ward, A.J., 2008. Consensus decision making by fish. *Current Biology* 18, 1773–1777.
- Surowiecki, J., 2005. *The Wisdom of Crowds*. Random House, Inc.
- Tolhurst, D.J., Movshon, J.A., Dean, A.F., 1983. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research* 23, 775–785.
- Touryan, J., Gibson, L., Horne, J.H., Weber, P., 2011. Frontiers: real-time measurement of face recognition in rapid serial visual presentation. *Frontiers in Perception Science* 2, 1–8.
- Troje, N., Bühlhoff, H.H., 1996. Face recognition under varying poses: the role of texture and shape. *Vis Research* 36, 1761–1771.