NeuroImage 106 (2015) 134–143

Contents lists available at ScienceDirect

NeuroImage

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

Neural representations of confidence emerge from the process of decision formation during perceptual choices

Sabina Gherman, Marios G. Philiastides *

Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, G12 8QB, UK
Centre for Cognitive Neuroimaging, University of Glasgow, 58 Hillhead Street, Glasgow, G12 8QB, UK

ARTICLE INFO

Article history:
Accepted 17 November 2014
Available online 22 November 2014

Keywords:
Decision making
Evidence accumulation
Confidence
Single-trial
EEG
Linear discriminant analysis

ABSTRACT

Choice confidence represents the degree of belief that one's actions are likely to be correct or rewarding and plays a critical role in optimizing our decisions. Despite progress in understanding the neurobiology of human perceptual decision-making, little is known about the representation of confidence. Importantly, it remains unclear whether confidence forms an integral part of the decision process itself or represents a purely post-decisional signal. To address this issue we employed a paradigm whereby on some trials, prior to indicating their decision, participants could opt-out of the task for a small but certain reward. This manipulation captured participants' confidence on individual trials and allowed us to discriminate between electroencephalographic signals associated with certain-vs.-uncertain trials. Discrimination increased gradually and peaked well before participants indicated their choice. These signals exhibited a temporal profile consistent with a process of evidence accumulation, culminating at time of peak discrimination. Moreover, trial-by-trial fluctuations in the accumulation rate of nominally identical stimuli were predictive of participants' likelihood to opt-out of the task, suggesting that confidence emerges from the decision process itself and is computed continuously as the process unfolds. Correspondingly, source reconstruction placed these signals in regions previously implicated in decision making, within the prefrontal and parietal cortices. Crucially, control analyses ensured that these results could not be explained by stimulus difficulty, lapses in attention or decision accuracy.

Introduction

Imagine running in the park on a rainy day, trying to discern whether the person across the lawn is an old friend. The decision to keep concentrating on your stride or change directions to go greet them depends on your level of confidence that it is really them. Choice confidence is crucial not only for such mundane tasks, but also for more biologically and socially complex situations. It provides a probabilistic assessment of expected outcome and can play a key role in how we adjust in ever-changing environments, learn from trial and error, make better predictions, and plan future actions.

In recent years, systems and cognitive neuroscience started to examine the neural correlates underlying perceptual decision making. As a result, many monkey neurophysiology (Gold and Shadlen, 2007; Kim and Shadlen, 1999; Mazurek et al., 2003; Newsome et al., 1989; Shadlen et al., 1996; Shadlen and Newsome, 2001), human neuroimaging (Cheadle et al., 2014; Heekeren et al., 2004, 2006, 2008; Ho et al., 2009; Ploran et al., 2007; Tosoni et al., 2008), and human electrophysiology (de Lange et al., 2010; Donner et al., 2007, 2009; O'Connell et al., 2012; Philiastides et al., 2006; Philiastides and Sajda, 2006; Ratcliff et al., 2009; Wyart et al., 2012) experiments provided converging support that perceptual decisions are characterized by a noisy temporal accumulation of sensory evidence which culminates when an observer commits to a choice. Despite this progress, however, it remains unclear how confidence is represented in the human brain and what its relationship is with the decision process itself.

Current theoretical and experimental accounts have regarded confidence as a meta-cognitive event (i.e. an epiphenomenon of the decision process) that relies on new information arriving beyond the decision point (Fleming et al., 2012; Pleskac and Busemeyer, 2010; Yeung and Summerfield, 2012). Conversely, little has been done in the way of exploring whether confidence might emerge earlier in the decision process and before one commits to a choice. Evidence for the latter has recently emerged from a limited number of animal studies (Kepecs et al., 2008; Kiani and Shadlen, 2009; Shadlen and Kiani, 2013), proposing that choice confidence in perceptual judgments might be an inherent property of the decision process itself and that the same neural generators involved in evidence accumulation also encode choice confidence. To date, it remains unclear whether confidence forms an integral part of the decision process itself and whether it emerges from the same neural generators involved in accumulating evidence for the decision. Similarly, it is unknown whether confidence is reflected in the rate of evidence accumulation itself.

* Corresponding author at: 58 Hillhead Street, Glasgow, G12 8QB, UK.
E-mail address: Marios.Philiastides@glasgow.ac.uk (M.G. Philiastides).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.036
1053-8119/

To address these open questions, we collected electroencephalography (EEG) data during a binary, delayed-response, task in which correct responses were rewarded with monetary incentives. Importantly, on a random half of trials and after forming a decision, participants were given the option to opt out of the task for a smaller but sure reward (a form of post-decision wager (Kiani and Shadlen, 2009)). We expected participants to waive the sure reward when they were certain of their choice, and select it otherwise. This in turn allowed us to use a multivariate single-trial classifier to discriminate between certain-vs.-uncertain trials to identify the temporal characteristics of the neural correlates of choice confidence. Importantly, additional control analyses were carried out to ensure that confidence-related effects could not be explained by stimulus difficulty or trial-by-trial changes in attention.

Materials and methods

Participants

Nineteen subjects (7 males) aged between 18 and 36 years (mean = 23.4 years) participated in the experiment. All had normal or corrected-to-normal vision and reported no history of neurological problems. Written informed consent was obtained in accordance with the School of Psychology Ethics Committee at the University of Nottingham.

Stimuli and task

Stimuli consisted of 20 face (face database, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany) (Troje and Bulthoff, 1996) and 20 car greyscale images obtained from the Web (size 500 × 500 pixels, 8-bits/pixel). Spatial frequency, contrast, and luminance were equalized across all images, and the magnitude spectrum of each image was adjusted to the average magnitude spectrum of all images. We manipulated the phase spectrum of the images to obtain noisy stimuli of varying levels of sensory evidence (i.e. we manipulated the percentage phase coherence of our images) (Dakin et al., 2002). Stimuli were presented centrally on a plain grey background on a computer screen using PsychoPy software (Peirce, 2007). The display was situated 1 m away from the subject, with each stimulus subtending approximately 8 × 8 degrees of visual angle.

We used a training session prior to the main task to identify subject-specific phase coherence values for the stimuli used in the main task. Specifically, during training subjects were required to perform a simple speeded face vs. car categorizations over a total of 600 trials, using images with 7 different phase coherence values (27.5–42.5%, in increments of 2.5%). Each image was presented for 0.1 s, and subjects were allowed a maximum of 1.25 s to make a response. The response was followed by an inter-trial interval, randomized between .75 and 1.5 s. There were an equal number of face and car stimuli, and these were presented in random order. Based on performance during this session, we selected three subject-specific phase coherence levels for the main task (henceforth referred to as low, medium, and high), which spanned psychophysical threshold (in the range 60–80% accuracy).

For the main experiment, subjects performed face vs. car categorizations during a delayed-response, post-decision wagering paradigm designed to discriminate between certain and uncertain trials (Fig. 1A). Importantly, on a random half of the trials, subjects were offered the option to opt-out of the task for a smaller (relative to a correct response) but sure reward (SR). This manipulation encouraged subjects to select the SR option on low confidence trials (Kiani and Shadlen, 2009). Responses were rewarded with points (correct = 10 points, incorrect = 0 points, SR choice = 8 points). The total number of points collected was translated into a monetary payment at the end of the experiment. Each trial began with a face or car stimulus presented for 0.1 s at one of the three possible sensory evidence levels. Stimulus presentation was followed by a forced delay (i.e., the decision time) randomized between 0.9 and 1.4 s. This delay was introduced prior to revealing whether participants could opt-out of the task, to ensure they formed a decision on every trial. Next, a visual response cue (1 s) informed participants whether or not the SR option would be available—this was indicated by a green or red fixation cross, respectively. In addition, the letters "F" (for face) and "C" (for car) where positioned randomly to the left and right of the central fixation cross to indicate the mapping between stimulus and motor effectors (right index and ring fingers, respectively). The latter manipulation aimed at separating the decision process from motor planning and execution. Subjects indicated their choice by pressing one of three buttons on a response box (LEFT/RIGHT for a stimulus choice, MIDDLE for the SR). They were instructed to respond after the response cue was removed from the screen. A response was followed by an inter-trial interval randomized in the range 1–1.5 s. Overall subjects performed 480 trials, divided into two blocks of 240 trials each.

EEG data acquisition

We recorded EEG data during performance of the main task, in an electrostatically shielded room, using a DBPA-1 digital amplifier (Sensorium Inc., VT, USA), at a sampling rate of 1000Hz. We used 117 Ag/AgCl scalp electrodes and three periocular electrodes placed below the left eye and at the left and right outer canthi. Additionally, a chin electrode was used as ground. All channels were referenced to the left mastoid. Input impedance was adjusted to <50 kOhm. To obtain accurate event onset times we placed a photodiode on the monitor to detect the onset of the stimuli. An external response device was used to collect response times. Both signals were collected on two external channels on the EEG amplifiers to ensure synchronization with the EEG data.

EEG data pre-processing

We applied a 0.5–100 Hz band-pass filter to the data to remove slow DC drifts and high frequency noise. These filters were applied non-causally (using MATLAB "filtfilt") to avoid phase related distortions. Additionally, we re-referenced our data to the average of all electrodes. To remove eye movement artifacts, participants performed an eye movement calibration task prior to the main experiment, during which they were instructed to blink repeatedly several times while a central fixation cross was displayed in the center of the computer screen, and to make lateral and vertical saccades according to the position of the fixation cross. We recorded the timing of these visual cues and used principal component analysis to identify linear components associated with blinks and saccades, which were then removed from the EEG data (Parra et al., 2005). Finally, we baseline corrected our EEG data, with the baseline interval defined as the 100 ms prior to stimulus onset.

Single trial EEG analysis

To identify confidence-related activity in the neural data, we used a single-trial multivariate discriminant analysis (Parra et al., 2002, 2005) to estimate linear spatial weightings of the EEG sensors, which discriminated between certain (SR waived) and uncertain (SR selected) trials. We applied our technique to discriminate between the two groups of trials at various time points, in the time range between 100 ms prior to, and 1000 ms following the presentation of the visual stimulus (i.e. during the decision phase of the trial). For each participant we estimated, within short pre-defined time windows of interest, a projection in the multidimensional EEG space (i.e. a spatial filter) that maximally discriminated between the two conditions on stimulus-locked data (Eq. (1)). Unlike conventional, univariate, trial-average event-related potential analysis, our multivariate approach is designed to spatially integrate information across the multidimensional sensor space, rather