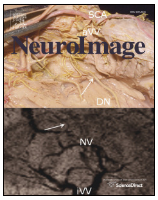
NeuroImage 109 (2015) 35–49

Contents lists available at ScienceDirect

NeuroImage

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



Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making

Joshua J. Tremel [a,b,c,*], Mark E. Wheeler [a,b,c,d,e,**]

[a] Department of Psychology, University of Pittsburgh, PA, USA
[b] Learning Research and Development Center, University of Pittsburgh, PA, USA
[c] Center for the Neural Basis of Cognition, University of Pittsburgh, PA, USA
[d] Center for Neuroscience, University of Pittsburgh, PA, USA
[e] School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

ARTICLE INFO

Article history:
Accepted 28 December 2014
Available online 3 January 2015

Keywords:
perceptual decision-making
evidence accumulation
diffusion model
fMRI
decision time

ABSTRACT

During a perceptual decision, neuronal activity can change as a function of time-integrated evidence. Such neurons may serve as decision variables, signaling a choice when activity reaches a boundary. Because the signals occur on a millisecond timescale, translating to human decision-making using functional neuroimaging has been challenging. Previous neuroimaging work in humans has identified patterns of neural activity consistent with an accumulation account. However, the degree to which the accumulating neuroimaging signals reflect specific sources of perceptual evidence is unknown. Using an extended face/house discrimination task in conjunction with cognitive modeling, we tested whether accumulation signals, as measured using functional magnetic resonance imaging (fMRI), are stimulus-specific. Accumulation signals were defined as a change in the slope of the rising edge of activation corresponding with response time (RT), with higher slopes associated with faster RTs. Consistent with an accumulation account, fMRI activity in face- and house-selective regions in the inferior temporal cortex increased at a rate proportional to decision time in favor of the preferred stimulus. This finding indicates that stimulus-specific regions perform an evidence integrative function during goal-directed behavior and that different sources of evidence accumulate separately. We also assessed the decision-related function of other regions throughout the brain and found that several regions were consistent with classifications from prior work, suggesting a degree of domain generality in decision processing. Taken together, these results provide support for an integration-to-boundary decision mechanism and highlight possible roles of both domain-specific and domain-general regions in decision evidence evaluation.

Introduction

Models of perceptual choice characterize decisions as processes in which evidence accumulates in a decision variable toward a boundary, and a choice is made when this boundary is reached. Electrophysiological studies have identified neuronal firing rate patterns resembling an accumulation-to-boundary mechanism in a number of non-human primate brain regions, including the superior colliculus (Ratcliff, 2003; Ratcliff et al., 2007), lateral intraparietal area (Shadlen and Newsome, 2001), dorsolateral prefrontal cortex (Kim and Shadlen, 1999), and frontal eye fields (Hanes and Schall, 1996). These regions show time series of neuronal activity consistent with predictions of sequential sampling models (Ratcliff, 1978; Usher and McClelland, 2001), in which the rate of accumulated neural activity is related to the time to make a decision (Gold and Shadlen, 2001; Hanes and Schall, 1996). Activity in these neurons is influenced by both the quality and availability of sensory evidence and thus may reflect a decision variable (Ditterich et al., 2003). Similar effects have been found in humans. Johnson and Olshausen (2003) used event-related potentials (ERP) and found that the voltage in the mid-frontal (FZ) electrode changed most rapidly for the fastest decisions. In a perceptual discrimination task, Philiastides and Sajda (2006, 2007) identified a late ERP component (~300 ms) that tracked aspects of the mean drift-rate in Ratcliff's drift-diffusion model (DDM). Using whole-brain fMRI, Heekeren et al. (2004) found a left superior frontal region whose peak activity reflected the strength of evidence for face-versus-house categorization (see also: Huettel et al., 2005; Kayser et al., 2010; Tosoni et al., 2008).

Due to limitations of EEG and fMRI, it is difficult to localize the source of time-dependent signals. However, by limiting the rate of information, the timescale of a decision can be slowed in order to compensate for the

---

* Correspondence to: J.J. Tremel, University of Pittsburgh, 647 Learning Research and Development Center, 3939 O'hara Street, Pittsburgh, PA 15260.
** Correspondence to: M.E. Wheeler, Georgia Institute of Technology, School of Psychology, 654 Cherry Street, Atlanta, Georgia, 30332-0170.
E-mail addresses: j.tremel@pitt.edu (J.J. Tremel), mark.wheeler@psych.gatech.edu (M.E. Wheeler).

http://dx.doi.org/10.1016/j.neuroimage.2014.12.072
1053-8119/

low temporal resolution of fMRI (Bowman et al., 2012; Carlson et al., 2006; Gluth et al., 2012; James et al., 2000). For example, in an object identification task, Ploran et al. (2007) gradually revealed objects over 16 seconds and found that activity in thirteen regions accumulated at a rate correlating with decision time. The role of some of these regions as evidence accumulators is supported by other observations. First, the rate of fMRI accumulation is influenced by errors—fMRI activity increases faster prior to the decision when an object is incorrectly identified compared to correct trials (Wheeler et al., 2008). Second, the rate and magnitude of accumulation is significantly less when people fail to commit to a decision (Ploran et al., 2011). Despite this support, it remains unclear whether this activity reflects the integration of information, or instead is due to hemodynamic artifacts or epiphenomenal cognitive events, such as attention, time-on-task, or urgency.

If fMRI accumulation reflects evidence accumulation, it should be directly related to evidence sources, and thus content-specific. To test this prediction, an extended face-house discrimination task was used to examine the evolution of activity within face- and house-selective regions in inferior temporal (IT) cortex (Haxby et al., 1994; Kanwisher et al., 1997). Behavioral and neuroimaging analysis were employed together with drift-diffusion models to directly link model hypotheses to task data. In the task, subjects viewed dynamic movies of noise-degraded faces and houses and made discriminations when reasonably confident. Importantly, the aim of this study differs from previous studies. For example, Heekeren et al. (2004) identified putative regions that appear to compute a decision-rule ("comparators"), but did not directly test whether temporally dynamic changes in IT activity are related to the decision process. Our primary aim was to localize face- and house-selective regions in IT and test whether activity follows an accumulation-like pattern in a content-specific manner. The use of dynamic movies with limited sensory information allowed us to increase the variance of decision times across many seconds, facilitating the investigation of time-sensitive effects using fMRI.

Materials and Methods

Subjects

Twenty-two healthy, right-handed, native English speakers with normal or corrected-to-normal vision participated in a 1.5-hour behavioral and functional magnetic resonance imaging (fMRI) session. Six subjects in total were excluded for excessive movement during scanning ($N = 4$), incomplete scan data ($N = 1$), or insufficient behavioral data in all conditions for reliable analysis ($N = 1$). The remaining 16 subjects (10 female) ranged in age from 21 to 26 years (mean 23.3). Informed consent was obtained from all subjects according to procedures approved by the University of Pittsburgh Institutional Review Board. Subjects were compensated $75 for their time.

Task

Subjects participated in a functional imaging scan and performed a face/house discrimination task. Subjects viewed short videos of noise-degraded faces and houses and made a forced-choice face/house decision. Stimuli were presented in a 300 × 300 pixel frame centered on a black background and were projected onto a screen at the head of the magnet bore at 1024 × 768 resolution. Subjects viewed the task via a mirror mounted to the radio frequency coil. Subjects were instructed to make a face or house decision when they reached a reasonable level of confidence and indicated their choice with an index-finger button press. Face and house responses were mapped to opposite hands, counterbalanced across subjects. MRI-safe projection equipment and fiber optic response glove system were produced by Psychology Software Tools (PST, Pittsburgh, PA). The PsychoPy software package was used for task presentation and data collection (Peirce, 2007, 2008). Response times (RTs) were recorded and used to approximate decision latency.

The task implemented a widely spaced event-related design (Fig. 1a). Face and house videos were displayed for 6 s, followed by a 10.5 s inter-trial interval (ITI) period to allow the blood-oxygen-level-dependent (BOLD) signal to approach basal levels before the next trial. Trials were further separated with additional ITI jitter of variable length, sampled randomly from an distribution of 0-6 s positively skewed toward the shorter intervals (1.5 s increments, mean 1.64 s) (Dale, 1999). Fully noise-degraded stimuli (~100% noise, completely randomized phase matrix) were displayed during fixation and ITI periods. Each trial period was indicated to the subject with a colored border (4 pixel width) surrounding the stimulus display frame. A green border indicated the trial period; the border turned grey for long ITI periods and red for jitter periods.

Testing took place over seven runs of 32 trials per run. Each run consisted of 14 face trials, 14 house trials, and four fully noise-degraded trials (two runs featured one extra face and house trial each, and two fewer full-noise trials). Noise levels were balanced across the entire session, featuring 17 trials per face/house at 65-69% noise, 15 at 70% noise, and 24 fully degraded images. Given our limited stimulus set, most stimuli were repeated twice per subject (with 16 stimuli per face/house shown 3 times); repeats of the same stimulus, however, were shown at noise levels differing by at least 2%. Behavioral analysis indicated no performance increase due to repetition (linear trend contrast of stimulus repetition, $F_{(1,\ 15)} = 1.48$, $p = 0.24$). Each subject received a stimulus set with a randomized presentation sequence and unique distribution of noise levels across stimuli.

Stimuli

The stimulus set consisted of 42 neutral-expression, frontal-view face and 43 house greyscale images, transformed into 6-second movie clips. The source images measured 512 × 512 pixels. Face stimuli were a part of the MacBrain Face Stimulus Set (courtesy of the MacArthur Foundation Research Network on Early Experience and Brain Development, Boston, MA). House images were compiled from searches for public domain photos (Google images) and from photographs around the Pittsburgh area. Backgrounds in the images were erased and cropped, leaving only the face or house on a white background. The stimulus set was entered into the following processing routine to normalize images across the set and introduce graded amounts of noise into each image. Using Matlab (2010a, The Mathworks Inc., Natick, MA), this routine computed the two-dimensional, forward discrete Fourier transform (DFT) of each image using a fast Fourier transform (FFT) algorithm. Each image's DFT was decomposed into a phase angle matrix and an amplitude matrix. The amplitude matrices of all images in the set were averaged, to help balance differences in image properties across the set, such as contrast, luminance, and brightness. Each individual phase matrix was convolved with a matrix of random noise using an additive white Gaussian noise (AWGN) filter, which combines signal (face or house phase matrix) and noise at a specified signal-to-noise ratio (SNR), producing a single, noise-convolved phase matrix. This allowed us to parametrically vary the amount of discrete evidence available for a given stimulus. To produce the final image, the average amplitude matrix was recombined with an individual noise-convolved phase matrix, and inverted with an inverse FFT algorithm (Heekeren et al., 2004).

Because the AWGN filter produced a random spatial distribution (but constant amount) of noise across an image, this procedure was repeated 90 times for each stimulus, generating 90 still images of the same face or house with identical SNR, but different patterns of noise. These frames were concatenated to produce a 6-second movie clip (at 15 frames per second). Because of the length and low SNR of the movies, the normally sub-second face/house decision process was effectively extended to several seconds, while allowing parametric