Journal of Neuroscience Methods 281 (2017) 21–32

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Research article

Evidence accumulation detected in BOLD signal using slow perceptual decision making

Paul M. Krueger[a,*], Marieke K. van Vugt[a,1], Patrick Simen[a,2], Leigh Nystrom[a], Philip Holmes[b], Jonathan D. Cohen[a]

[a] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, 08544, USA
[b] Department of Mechanical and Aerospace Engineering and Program in Applied and Computational Mathematics; Associated Faculty in Neuroscience, Princeton University, Princeton, NJ, 08544, USA

HIGHLIGHTS

- Measuring evidence accumulation with fMRI is challenging because the BOLD signal is slow.
- We protracted evidence accumulation using predictions of the drift-diffusion model.
- We minimized collinearity of regressors to distinguish ramping and "boxcar" activity.
- Results include new candidate regions for evidence accumulation and exclude parietal cortex.

ARTICLE INFO

Article history:
Received 4 August 2016
Received in revised form 19 January 2017
Accepted 22 January 2017
Available online 25 January 2017

Keywords:
Evidence accumulation
Drift-diffusion model
Decision making
Functional magnetic resonance imaging

ABSTRACT

Background: We assessed whether evidence accumulation could be observed in the BOLD signal during perceptual decision making. This presents a challenge since the hemodynamic response is slow, while perceptual decisions are typically fast.
New method: Guided by theoretical predictions of the drift diffusion model, we slowed down decisions by penalizing participants for incorrect responses. Second, we distinguished BOLD activity related to stimulus detection (modeled using a boxcar) from activity related to integration (modeled using a ramp) by minimizing the collinearity of GLM regressors. This was achieved by dissecting a boxcar into its two most orthogonal components: an "up-ramp" and a "down-ramp." Third, we used a control condition in which stimuli and responses were similar to the experimental condition, but that did not engage evidence accumulation of the stimuli.
Results: The results revealed an absence of areas in parietal cortex that have been proposed to drive perceptual decision making but have recently come into question; and newly identified regions that are candidates for involvement in evidence accumulation.
Comparison with existing methods: Previous fMRI studies have either used fast perceptual decision making, which precludes the measurement of evidence accumulation, or slowed down responses by gradually revealing stimuli. The latter approach confounds perceptual detection with evidence accumulation because accumulation is constrained by perceptual input.
Conclusions: We slowed down the decision making process itself while leaving perceptual information intact. This provided a more sensitive and selective observation of brain regions associated with the evidence accumulation processes underlying perceptual decision making than previous methods.

* Corresponding author. Current address: Department of Psychology, University of California, Berkeley, Berkeley, CA, 94720, USA.
E-mail addresses: paul.m.krueger@berkeley.edu (P.M. Krueger), m.k.van.vugt@rug.nl (M.K. van Vugt), psimen@oberlin.edu (P. Simen), nystrom@princeton.edu (L. Nystrom), pholmes@math.princeton.edu (P. Holmes), jdc@princeton.edu (J.D. Cohen).
[1] Current address: Department of Artificial Intelligence and Cognitive Engineering, University of Groningen, Groningen, Netherlands.
[2] Current address: Department of Neuroscience, Oberlin College, Oberlin, OH, 44074, USA.

http://dx.doi.org/10.1016/j.jneumeth.2017.01.012
0165-0270/

1. Introduction

The neural correlates of perceptual decision making have been studied extensively using monkey neurophysiology (e.g., Gold and Shadlen, 2000) and human electroencephalography (EEG; e.g., van Vugt et al., 2012), magnetoencephalography (MEG; e.g., Donner et al., 2009), and functional magnetic resonance imaging (fMRI; e.g., Heekeren et al., 2006; Wheeler et al., 2015; see Mulder et al., 2014, for a meta analysis). In comparison to the other methods, the temporal resolution of fMRI (on the order of 2 s) is very slow—too slow, in fact, to capture the sub-second time course of most perceptual decisions. Nevertheless, it has been the neuroimaging method of choice used to distinguish which brain regions are involved in perceptual decisions in humans (e.g. Heekeren et al., 2006; Ploran et al., 2007; Ivanoff et al., 2008; Tosoni et al., 2008; Basten et al., 2010; Ploran et al., 2011; Gluth et al., 2012).

The dynamics of the decision making process are typically described by models of evidence accumulation such as the drift diffusion model (DDM; Ratcliff, 1978), linear ballistic accumulator (Brown and Heathcote, 2008), or leaky competing accumulators (Usher and McClelland, 2001). These models posit that in order to make simple decisions, people accumulate noisy evidence supporting competing alternatives until the evidence in favor of one reaches a decision threshold. The noisiness of the evidence and the amount of drift favoring a particular alternative represent the quality of the signal on which the decision is based, or the difficulty of the task. The threshold is used to model a participant's speed-accuracy tradeoff, with higher thresholds representing more cautious decisions.

There have been several approaches to investigating the neural correlates of evidence accumulation using fMRI. Some have argued that neural correlates of evidence accumulation should be effector-specific, while others have argued the opposite. In the first camp, Tosoni et al. (2008) found increases in BOLD activity in precuneus, medial parietal lobe, posterior parietal reach region, precentral gyrus and sensorimotor cortex for button press responses, whereas saccade responses elicited more activation in the posterior intra-parietal sulcus and the frontal eye field. Only a subset of these areas showed activation that was modulated by the amount of sensory evidence that was accumulated. Other studies have focused on non-effector-specific responses. For example, Heekeren et al. (2006) found that a large number of areas including the dorsolateral prefrontal cortex (dlPFC) and inferior parietal lobe (IPL) showed activity consistent with evidence accumulation. In a study of value-based decision making, Basten et al. (2010) used individual differences in each participant's drift rate in addition to modulation by task difficulty to identify clusters in intraparietal sulcus as neural correlates of evidence accumulation. A problem with these approaches is that due to the slow nature of fMRI blood-oxygenation-level-dependent (BOLD) activity, it confounds evidence accumulation with several other processes that occur simultaneously, such as perceptual processing and motor preparation.

Another approach to studying evidence accumulation has been to slow down the decision such that it can be tracked even with the sluggish BOLD response. For example, Ploran et al. (2007) and Ploran et al. (2011) gradually revealed perceptual evidence used by participants to determine object identity. They found that BOLD activity in inferior frontal, temporal, and parietal regions showed a progressive increase in agreement with a model of gradual accumulation of evidence. Using a similar discontinuous stream of evidence, Gluth et al. (2012) showed that in a value-based decision task, the supplementary motor area, caudate nucleus and anterior insula tracked the amount of collected evidence. Ivanoff et al. (2008) also slowed down responses by gradually increasing the amount of visible evidence in a random dot motion task over the course of each trial. They examined the dynamics of evidence accumulation in a set of ROIs obtained from an easy motion discrimination localizer task. They found that in lateral prefrontal cortex (PFC) and supplementary motor area (SMA) activity increased with accumulated evidence, and this region also discriminated between speed- and accuracy conditions.

All of these paradigms are potentially problematic, however, in that they slowed the accumulation process by manipulating the availability of evidence (i.e., slowly strengthening the stimulus or gradually revealing it). This progressive increase in the availability of evidence could have induced a progressive increase in the activity of regions responsible for visual perception, without reflecting the operation of an integration process at later cognitive stages (as discussed in more detail below). Thus, these designs introduce a potential confound between perceptual detection and integration because decision making is constrained by the availability of perceptual information rather than the accumulation of already-available information. This makes it difficult to dissociate these two subprocesses of decision making.

Here, we suggest an alternative approach to observing the evidence accumulation process in fMRI, based on a theoretical prediction from the DDM. The DDM predicts that it should possible to dramatically slow down the decision process itself — without interfering with stimulus input — by modifying the speed-accuracy trade-off. As suggested by both Bogacz et al. (2006) and Balci et al. (2011), if participants are seeking to maximize their reward rates and therefore their total earnings, optimality analysis of decision behavior reveals that it should be possible to slow down responses by imposing penalties for incorrect responses (we provide an explanation for this effect in greater detail below). This offers a more ecologically valid way to protract evidence accumulation, without manipulating or otherwise impacting the availability of evidence from the stimulus.

We conducted an experiment to test whether we could slow down the decision process in this way, and use this to measure the neural correlates of the integration component. Participants performed a random dot motion task (Britten et al., 1992), in which every correct answer was rewarded with one cent, while they incurred a penalty of two cents for every incorrect response. This resulted in a dramatic increase of response times, making it possible to disentangle signals related to different subprocesses in decision making using fMRI. We compared regressors designed to distinguish ramping activity from "boxcar" activity (van Vugt et al., 2012). By further comparing activity between the decision task and a control task that did not engage a perceptual evidence accumulation process, we identified brain regions involved in perceptual evidence accumulation unconfounded by stimulus factors, that included bilateral anterior insula, cerebellum, caudate, frontal gyrus, and higher visual areas.

2. Materials and methods

2.1. Motion discrimination task

Participants engaged in a visual decision making task in which they judged the net direction (leftward or rightward) of a cloud of randomly moving dots (Fig. 1A), indicating responses bimanually with a button box. Each trial began with the onset of moving dots, and participants were free to respond at any time. After making a decision, participants were given feedback indicating whether they answered correctly or not, and the running total number of points earned.

The random dot kinematograms were similar to those used in a series of psychophysical and decision making experiments involving monkeys (e.g., Britten et al., 1992; Gold and Shadlen, 2001;

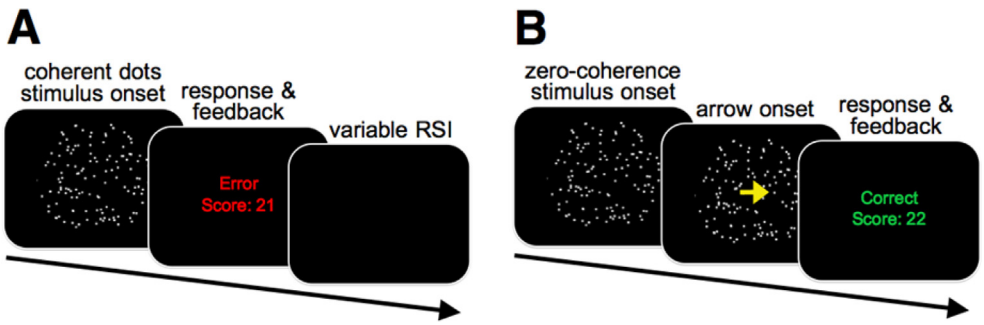

Fig. 1. Moving dots task. (A): A single trial from the experimental condition. After a brief fixation cross, a circular patch of moving dots appeared onscreen. Most of the dots moved in random directions, but a small subset of dots moved coherently in a single direction (left or right). Participants freely responded with bimanual button presses to indicate which direction they thought the coherent dots were moving, at which point feedback appeared onscreen, indicating whether they were correct and the running total of points. (B): A trial from the control condition. Participants viewed randomly moving dots (with zero coherence) that they were told to ignore, and were then cued by an arrow to respond left or right. The onset of the arrow was timed such that the distribution of dots motion viewing times was the same as the experimental condition.

Shadlen and Newsome, 2001) and humans (e.g. Simen et al., 2009; Balci et al., 2011) as participants. Stimuli consisted of an aperture of an approximately 3 inch diameter viewed from approximately 100 cm (approximately 4° visual angle) in which white dots (2 × 2 pixels) moved on a black background. A subset of dots moved coherently either to the left or to the right on each trial, whereas the remainder of dots were distractors that jumped randomly from frame to frame. Motion coherence was defined as the percentage of coherently moving dots. Dot density was 17 dots/square degree, selected so that individual dots could not easily be tracked. The experiment presentation code was written in PsychToolbox (Brainard, 1997). Dots were presented with PsychToolbox extensions written by J. I. Gold (http://code.google.com/p/dotsx/).

We calibrated the motion coherence for each participant individually, so that they performed at approximately 75% and 85% correct for low and high coherence stimuli, respectively. Before the neuroimaging experiment, participants were trained for 1–4 behavioral sessions until they were familiar with the task, and their performance was consistently in the target range of 75% and 85% percent correct. Both behavioral-only and fMRI sessions consisted of three ~20-min runs of four 4.5-min blocks each, with a 30-s break between blocks. Participants could do as many trials as they desired in each block, but were encouraged to maximize the points they earned over the course of each block. Maximizing points therefore depended both on the number of trials performed and their accuracy. To prevent anticipatory responses and to facilitate deconvolution of the BOLD signal, the response-to-stimulus interval between each trial was drawn from a gamma distribution with a mean of ten seconds, a standard deviation of one second, and a skewness of 0.5.

2.2. Slowing down decision time

Participants were paid one cent for every correct response, but they lost two cents for every incorrect response. To ensure they would not end up with a negative balance, every participant was given a starting credit of $5. Theoretical analyses (Bogacz et al., 2006; Balci et al., 2011) predicted that introduction of this penalty should cause participants to slow down dramatically. If responses are slowed down on the order of several seconds, this should allow for analysis of accumulation dynamics, even after their convolution with the time course of the BOLD signal.

The theoretical analysis used to slow down response times is based on evidence that participants try to maximize their reward rate in tasks of this type (Balci et al., 2011; Bogacz et al., 2010; Simen et al., 2009; Starns and Ratcliff, 2010). Participants' reward rates are determined by a tradeoff between responding quickly and responding accurately. Without penalties imposed for errors, the expected reward rate (RR) is the proportion of correct responses divided by the average time between them (Bogacz et al., 2006):

$$RR = \frac{1 - ER}{DT + T_0 + RSI} \tag{1}$$

where *ER*, the error rate, is the proportion of incorrect responses, and the terms in the denominator sum up to the total time between trials (*DT*, the decision time, is the period of evidence accumulation; T_0, the non-decision time, is the fraction of the response time with no evidence accumulation, e.g., to execute a motor response; *RSI*, the response-to-stimulus interval, is the time between trials). The optimal decision time as a function of error rate is given by the DDM as:

$$\frac{DT}{D_{total}} = \left(\frac{1}{ERlog\frac{1-ER}{ER}} + \frac{1}{1-2ER} \right)^{-1} \tag{2}$$

where the term on the left is the normalized decision time ($D_{total} = T_0 + RSI$). If participants are penalized for incorrect responses, then the modified expected reward rate is (Bogacz et al., 2006):

$$RR_m = \frac{(1 - ER) - qER}{DT + D_{total}} \tag{3}$$

where *q* is the penalty for incorrect responses. The optimal normalized decision time becomes:

$$\frac{DT}{D_{total}} = (1 + q) \left(\frac{\frac{1}{ER} - \frac{q}{1-ER}}{log\frac{1-ER}{ER}} + \frac{1-q}{1-2ER} \right)^{-1} \tag{4}$$

Fig. 2 shows optimal performance curves produced by Eq. (4). When no penalties are imposed (thin blue curve), the optimal decision time is small relative to the total decision time (DT_{total}). When penalties for errors are larger than rewards for correct responses, the optimal decision time increases dramatically (thick purple and green curves), and becomes unbounded at error rates below 50% (Balci et al., 2011) (when the penalty/reward ratio is 2.0, optimal decision times become unbounded as error rates approach approx. 20.8% (green curve); when the ratio is 1.5, decision times become unbounded as error rates approach approx. 31.0% (purple curve)).

Whether the predicted, protracted process of evidence accumulation is likely to occur in the brain is an open question. An influential claim has been made that the DDM only accurately models decision processes lasting one second or less (e.g., Ratcliff and Rouder, 1998). Despite this interpretation, the DDM has been shown to give a compelling account of data with much longer response times (e.g., Simen et al., 2016). Nevertheless, Ratcliff and Rouder (1998) conjecture that, good fit or bad, response times longer than a second probably reflect final decisions consisting of sequences of failed micro-decisions. Under this conjecture, each

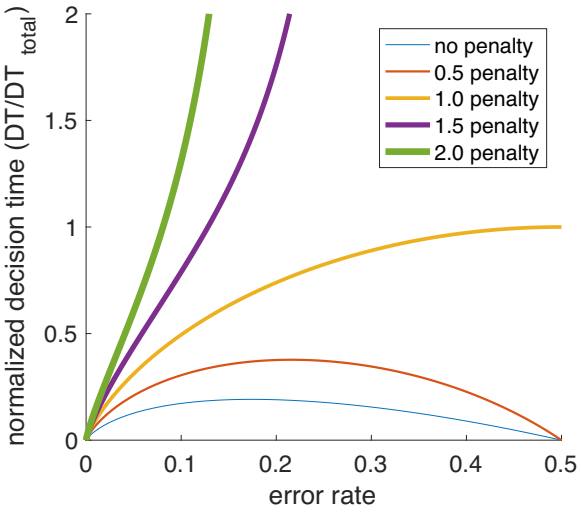

Fig. 2. Penalties for incorrect responses drastically increase the optimal decision time. Each curve shows the normalized decision time that maximizes reward rate, as a function of error rate, as predicted by the drift-diffusion model. When penalties for incorrect responses are small relative to rewards for correct responses (thin blue and red curves), then the optimal normalized decision time is always less than one. When penalties are large relative to rewards (thick purple and green curves), then the optimal decision time increases dramatically. This theoretical prediction was used to slow down decision making enough to measure with BOLD fMRI.

micro-decision fails to produce a final choice. Instead, the final decision is produced by aggregating micro-decisions or it is based solely on the outcome of the last micro-decision. In either case, accumulator regions would be expected to show sequences of ramps that reset to their starting points many times within each completed decision, contrary to our assumption of a single, shallow ramp. Note that micro-decision aggregation might itself produce a crucial accumulation process in difficult decisions such as the ones we investigate here. In any case, we demonstrate that assuming a single, protracted diffusion process for each decision leads to behavior and model fits with properties predicted by the reward-rate optimality hypothesis, which could easily have been disconfirmed by the data. Using this method to infer which brain areas are involved in evidence accumulation therefore seems as reasonable as modifying the stimulus to delay decision making as was done by Ploran et al. (2007) and Gluth et al. (2012).

2.3. Control task

We used a control task in which no evidence accumulation of moving dots could take place, but participants still viewed comparable visual input and gave responses as they did in the experimental condition (Fig. 1B). Each control trial started with random dot motion (with 0% coherence), followed by a yellow arrow (approximately 2×2 inches, located in the center of the screen) indicating the direction to which a participant should respond. The arrow onset time was calibrated such that the dot motion viewing times (measured from dots onset to button press) from these "arrows" trials mirrored the response times of the preceding motion discrimination block (with coherence greater than 0%). This task therefore controlled for low-level visual input, dot kinematogram viewing times, and motor responding, while eliminating evidence accumulation for coherent dot movement. During every 20-min run, each 4.5-min block of the motion discrimination task was followed by a block of the control task. For shorthand, we often refer to the control task as the "arrows" task, and the motion discrimination task as the "dots" task.

We use the term "dot motion viewing time," or "viewing time" for short, to refer to the duration from the onset of dots to button press. The term "response time" can equivalently be used for the dots task; however, for the arrows task, "response time" is inaccurate since participants are passively viewing dots, and responding to the onset of the arrow. Therefore, "dot motion viewing time" can accurately describe this duration in both tasks.

2.4. Participants

We recruited a total of 30 participants (13 females, mean age: 22.7, range: 17.5-38.4). 10 participants were removed from the experiment because they failed to finish the pre-fMRI behavioral training sessions (9) or fell asleep in the scanner (1). Participants were recruited from the Princeton community and paid $20 per hour, plus a performance bonus based on the number of points they acquired during the experimental task. The experiment was approved by the Institutional Review Board of Princeton University.

2.5. DDM model fits

We used the DMA toolbox (VandeKerckhove and Tuerlinckx, 2007; VandeKerckhove and Tuerlinckx, 2008) to fit the DDM to the behavioral data. We allowed only drift rate to vary between conditions, since our design only varied stimulus coherence which primarily affects drift rate (e.g., Balci et al., 2011; Ratcliff, 2002; van Vugt et al., 2012). This was the best-fitting model using a BIC test for 12/20 participants (when comparing the following models: only drift varies; drift and threshold vary; and drift, threshold and non-decision time vary with coherence condition). The remaining 8 participants were equally split between the model that included only the threshold, and the one that included both threshold and non-decision time. We removed all RTs longer than 25 s prior to fitting the DDM to the behavioral data. This outlier removal rule affected data from only a minority of the participants, with an average of 1.4 outlier trials per participant out of an average 60 trials total (around 2%). For the purpose of fitting the DDM, it is worth noting that our participants make errors, which allow us to constrain the parameter fits with both correct and error RT distributions. Despite the long RTs in our experiment, we were able to calibrate the error rate by manipulating the stimulus coherence, as described in Section 2.1.

2.6. Data acquisition and preprocessing

Scanning was performed on a 3 T Siemens (Munich, Germany) Allegra scanner. First an anatomical scan with voxel size $1 \times 0.5 \times 0.5$ mm was acquired using an MPRAGE sequence. Subsequently, whole brain, BOLD-weighted echo-planar images were acquired parallel to the anterior commissure-posterior commissure line with a repetition time of 2 s and echo time of 30 milliseconds while participants performed the task. These images consisted of $64 \times 64 \times 34$ voxels, $3 \times 3 \times 4$ mm in size. We used AFNI (Cox, 1996) to preprocess the fMRI data. Functional images were first aligned to correct for head movement during the scan. Slice timing correction was then performed using sine interpolation. Images were subsequently normalized to Montreal Neurological Institute coordinates (the TT_N27 template) and resampled at $3 \times 3 \times 3\ mm^3$ resolution. A Gaussian smoothing kernel of 4 mm full-width at half-maximum (FWHM) was applied to improve comparison across subjects, and finally, all time series were scaled to have a mean signal value of 100. Head movement estimates derived from the first realignment step were included as regressors in all analyses to help diminish the impact of any movement-related effects on the results.

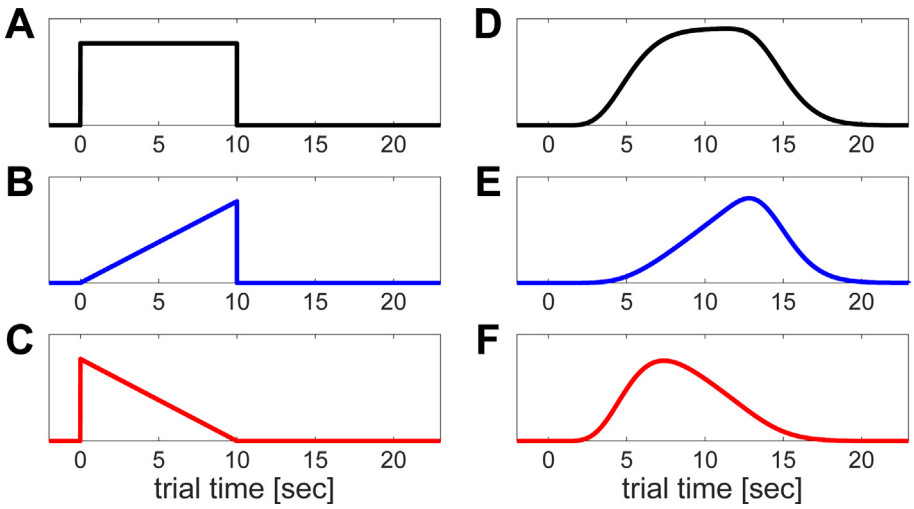

Fig. 3. Regressor types. The column on the left shows three types of regressors: (A) a boxcar function, (B) an up-ramp, and (C) a down-ramp. These example regressors are for a trial with a viewing time of 10 s. The column on the right (D, E, F) shows these same regressors after being convolved with the hemodynamic response function. We used only up-ramp and down-ramp regressors in order to minimize collinearity and thereby reliably distinguish ramping activity from boxcar activity.

2.7. General linear model

We performed a multiple linear regression that was designed to distinguish two component processes involved in perceptual decision making: detection and integration. A (motion) detector was defined as an area of the brain expected to be active at a more or less constant level throughout the decision process. We modeled such neural activity as a boxcar function that turned on at stimulus onset (i.e. onset of moving dots) and turned off at the button press (Fig. 3A). An integrator was defined as an area of the brain in which activation was expected to build up more or less constantly throughout the decision process when stimuli were presented continuously. We modeled such neural activity as a ramp that started from zero at stimulus onset and rose linearly until the button press, at which point it dropped back to zero (Fig. 3B). This assumption corresponds more closely to the linear ballistic accumulator (LBA) model of Brown and Heathcote (2008) than to the diffusion model, since it replaces the random-walk trajectory of a diffusion process with a straight line (see e.g., Ho et al., 2009). However, this assumption was necessary for our GLM analyses, since there is no behavioral signature of the diffusion process other than its termination time and the choice it selects. This assumption was partially justified by the low-pass-filter property of the BOLD signal, which would in any case tend to linearize a noisy process of ramping neural activity.

Ramp and boxcar regressors display collinearity, however, and the correlation between the two increases substantially when convolved with the low-pass hemodynamic response function (Figs. 3D & E). Using such regressors would limit the reliability of model fits and the interpretability of results because of the ambiguity between the two: areas of the brain that fit well to the ramp regressor may still largely display boxcar-like activity, and vice versa. In order to minimize the collinearity between regressors while still distinguishing detection and integration, we used two distinct regressors: an up-ramp, as before, and a non-up-ramp component of a boxcar, which looks like a "down-ramp": activity goes from zero to one at stimulus onset and ramps down to zero at the button press (Fig. 3C). Thus, the sum of the up-ramp regressor and the down-ramp regressor is exactly a boxcar (we thank Jack Grinband [personal communication] for suggesting this approach). Using this scheme, true ramping activation is detected where the goodness of fit of the up-ramp regressor is significantly greater than that of the down-ramp regressor; if the difference is not significant,

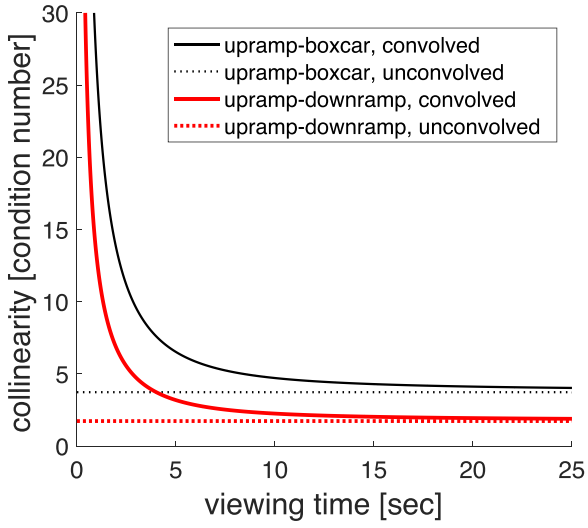

Fig. 4. Collinearity is minimized by using down-ramps instead of boxcar functions, and by having longer response times. The collinearity between a single unconvolved up-ramp (like that seen in Fig. 3B) and an unconvolved boxcar (as in Fig. 3A) is shown as a thin dotted black line. The collinearity between an unconvolved up-ramp and an unconvolved down-ramp (as in Fig. 3C) is lower (thick dotted red line). When these same regressors are convolved with the hemodynamic response function (as shown in the right column of Fig. 3), the collinearity increases. However, collinearity decreases as a function of viewing time. The thin solid black curve shows the collinearity between a single convolved up-ramp and a convolved boxcar. The thick solid red curve shows the collinearity between a convolved up-ramp and a convolved down-ramp, which is much lower.

we conclude that the activation does not unambiguously support the hypothesis of ramping. Thus, the down-ramp alone may not necessarily model BOLD activity, but when used in concert with an up-ramp improves our ability to distinguish upward-ramping activity from boxcar activity.

Fig. 4 illustrates that collinearity between regressors can be minimized by using up-ramp and down-ramp regressors instead of up-ramp and boxcar regressors, as well as by extending viewing times. Collinearity is measured using the condition number (Belshey et al., 1980), which in this case is the ratio of singular values of the matrix containing each pair of regressors. A smaller condition number indicates less collinearity. The condition number of a single up-ramp (as in Fig. 3B) and a single boxcar (as in

Fig. 3A) is 3.73, and does not vary depending on the length of the trial (thin dotted black line in Fig. 4). The condition number of an up-ramp and a down-ramp (as in Fig. 3C) is 1.73 (thick dotted red line in Fig. 4). When these regressors are convolved with the low-pass hemodynamic response function, the collinearity increases overall, but approaches that of the un-convolved regressors as the trial length increases (solid lines). The thin black solid curve shows the condition number of a convolved up-ramp (as in Fig. 3E) and a convolved boxcar (as in Fig. 3D). The thick red solid curve shows the condition number of a convolved up-ramp (as in Fig. 3E) and a convolved down-ramp (as in Fig. 3F). Note that the example regressors shown in Fig. 3 are for a trial with a viewing time of 10 s. With shorter viewing times, the discernibility of the convolved regressors diminishes.

Using down-ramp regressors instead of boxcars substantially decreases collinearity, thereby improving our ability to reliably detect ramping BOLD activity. Furthermore, extending viewing times has two advantages: 1) it decreases the collinearity of convolved regressors, further improving our ability to reliably detect ramping activity (Fig. 4); and 2) it makes it possible to measure ramping activity directly despite the low temporal resolution of fMRI (our data were acquired at 2 s per sample).

To account for the possibility that response-related activity may be mistaken for accumulation, we fit an additional GLM. In this GLM, we added a stick-function regressor at the time of responding to account for decision-related increases in BOLD that would be expected to occur near the time of responding, which could be mistaken for ramping. This analysis, however, made a negligible difference to any of our results. We also visually inspected the time-series data from each ROI, and found that the BOLD signal showed a pattern of ramping consistent with constant accumulation (Fig. S4).

In summary, our experimental paradigm had three key features that allowed us to measure perceptual evidence accumulation: 1) penalties for incorrect responses which dramatically slowed down response times and yielded relatively high accuracy despite low signal-to-noise ratios in the stimuli; 2) a control task with matched stimulus viewing times in which no dot-motion integration takes place, since the button press depends on the direction of a yellow arrow and not on dot motion; and 3) the use of regressors that allow us to measure ramping activation that is minimally confounded with boxcar activation. The underlying methodological challenge with measuring evidence accumulation using fMRI is the sluggishness of the BOLD response and the poor temporal resolution of fMRI. The combination of 1) & 3) allowed us to leverage an unprecedented degree of discrimination between ramping activation and boxcar activation. The use of 2) allowed us to also distinguish ramping activation that associated with evidence accumulation for the dot motion stimuli *per se* from that which is involved more generally during manual response preparation.

For each participant we fit the convolved up-ramp and down-ramp regressors to the voxel-wise BOLD fMRI data (with third-order polynomial baseline detrending on each twenty-minute scanner run) using multiple linear regression (AFNI's 3dDeconvolve). In total, our GLM had ten regressors for each participant: the up-ramp for the experimental motion discrimination ("dots") trials, the up-ramp for the control condition ("arrows") trials, the down-ramp for dots trials, the down-ramp for arrows trials, and six standard motion-correction regressors. To correct for multiple comparisons, we first estimated the spatial autocorrelation function (ACF) for each participant's residuals left over from the GLM (AFNI's 3dFWHMx). This ACF was fit to a mixture model of Gaussian plus mono-exponential. We then used the group mean ACF parameters to estimate the probability of false positives (using the most recent version of AFNI's 3dClustSim; Cox et al., 2016). We set a voxelwise threshold of $p < 0.001$ and a corrected threshold of $p < 0.01$ with bi-sided and face/edge nearest neighbor clustering parameters.

Using four parameter estimates (beta-weights) per participant from the GLM (up-ramp for dots condition, up-ramp for arrows condition, down-ramp for dots condition, and down-ramp for arrows condition), we performed a voxel-wise ANOVA with three factors: task type (dots condition vs. arrows condition), ramp regressor type (up-ramps vs. down-ramps), and participants (modeled as random effects; AFNI's 3dANOVA3). We identified voxels for which there was a significant main effect of condition on beta-weights, and an interaction between condition and ramp regressor type on beta-weights (each at $p < 0.001$, corrected). For areas involved in integration, the difference between beta weights for dots trials and arrows trials should be positive, and the interaction between trial type and ramp regressor type should be such that the difference between beta-weights for up-ramp regressors and down-ramp regressors is more positive for dots trials than for arrows trials. This would indicate significant upward ramping activity that is more significant for the decision making task than for the control task.

We also tested for voxels for which there was a significant increase in beta-weights from the arrows task to the dots task (measured using the difference in beta-weights between the dots task and arrows task, collapsed across ramp regressor type, at $p < 0.001$, corrected). These areas should be involved in the decision making process, although not necessarily involved in evidence accumulation. Within these areas, those involved in evidence accumulation should show a greater difference between beta weights for up-ramp regressors and down-ramp regressors for the dots task than for the arrows task, and overlap with regions found in the previous analysis. Areas for which there is no such interaction are assumed not to be strongly involved in evidence accumulation.

3. Results

3.1. Behavioral results

Before turning to the neuroimaging results, we first discuss effects related to the behavioral manipulations. Fig. 5A shows that participants had an average response time of 10.1 s for the low coherence condition and 7.2 s for the high coherence condition, indicating that we succeeded in slowing participants down dramatically from typical performance without penalties (on the order of 1 s). Participants had an average error rate of 26% for the low coherence condition, and 15% for the high coherence condition (Fig. 5B). The average coherence levels calibrated to each participant were 3.8% for low coherence and 7.6% for high coherence (Fig. 5C).

Fig. 6 visualizes the relation between dot motion viewing times in the experimental and control conditions, with each point reflecting the dot motion viewing time and the two-dimensional standard error of the mean for that participant. It is easy to see that dot motion viewing times were well-matched between the experimental condition (in which participants made perceptual decisions) and the control condition (in which participants passively viewed randomly moving dots).

We fit the DDM to each participant's behavioral data in order to assess whether this model could provide a good fit of the data in our task, and examine the model parameters. Table 1 shows the estimates of the drift parameter (note: the DMAT fitting algorithm sets the DDM noise parameter to 0.1, following conventions established by Ratcliff and colleagues). As expected, the drift rate for low coherence trials is lower than for high coherence trials [$t(19) = -6.03$, $p < 0.001$]. The thresholds do not differ between conditions because those were held fixed (see Methods). Note that these threshold values are extremely high, relative to typical DDM-fits in experiments without penalties (0.15 is common), and that drift values

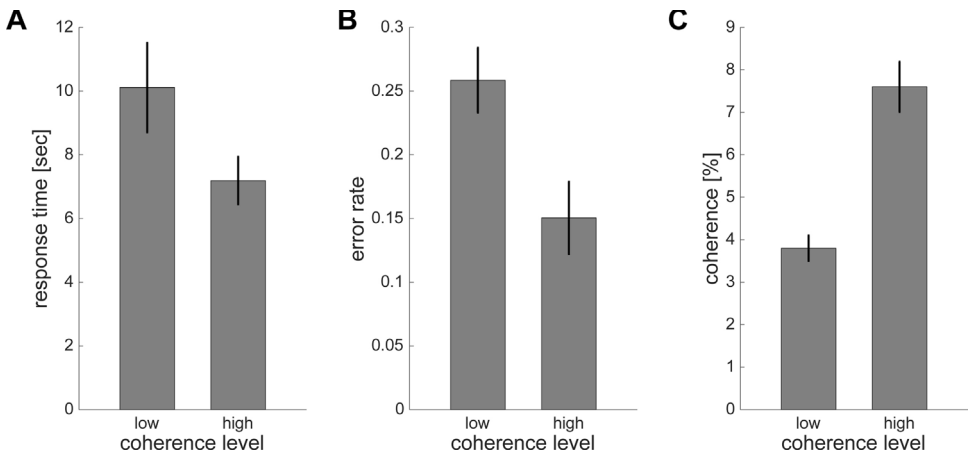

Fig. 5. Behavioral data for motion discrimination averaged across participants, with error-bars indicating the standard error of the mean. (A) response times for the high coherence level and the low coherence level (mean 10.1 and 7.2, respectively) and (B) the proportion of incorrect responses for high coherence low coherence trials (0.26 and 0.15, respectively). (C) The low coherence level and the high coherence level were calibrated for each participant to correspond to an error rate of about 25% and 15%, respectively. The average coherence (i.e. the percentage of dots moving in one direction) was 3.8% for the low coherence level and 7.6% for the high coherence level.

Table 1
Drift Diffusion Model parameter estimates. For each parameter, the mean and standard error across participants is shown for low coherence trials and high coherence trials. The very low estimates for non-decision times likely reflect the fact that estimates of these are generally in the sub-second range, which represents a substantially smaller fraction of the unusually long decision times in this dataset. Nevertheless, fits give quite a reasonable description of the response time distributions (see Fig. S1 for an assessment of model fit quality by means of expected versus observed quantile plots).

DDM parameter	Low coherence		High coherence	
	Mean	Standard error	Mean	Standard error
drift	0.016	0.002	0.036	0.005
threshold	0.57	0.02	0.57	0.02
non-decision time (sec)	0	0	0	0

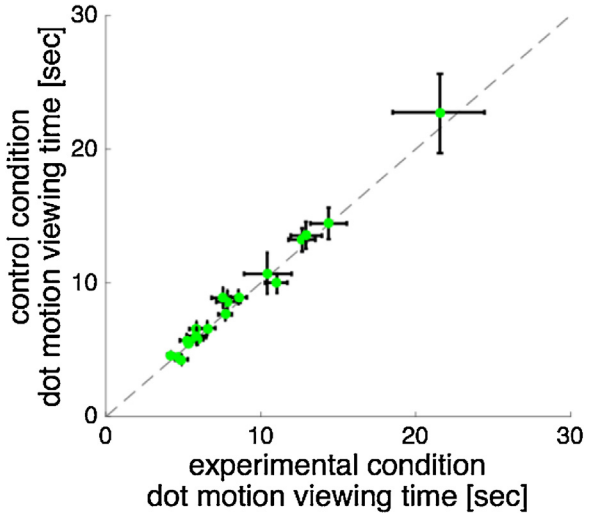

Fig. 6. Dot motion viewing times were matched between the control condition and the experimental condition. Each data point represents the mean dot motion viewing time for a single participant in the motion discrimination task and the control task, with error-bars indicating the standard error of the mean for that participant for the control condition (vertical error-bars) and the experimental condition (horizontal error-bars).

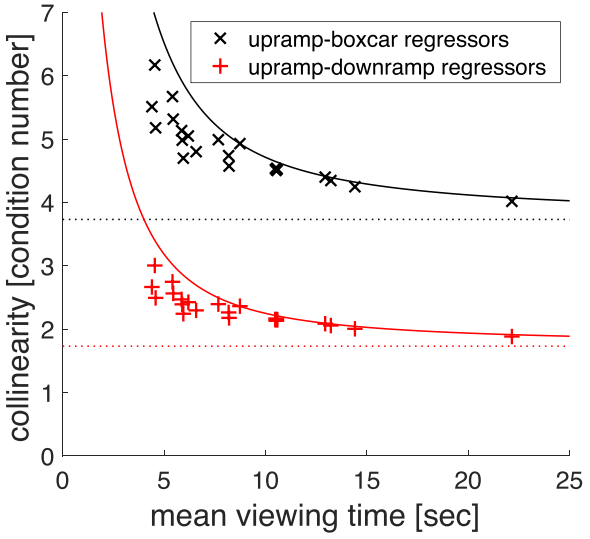

Fig. 7. Collinearity of actual regressors. Each cross shows the collinearity of regressors for one participant, plotted against each participant's mean dot motion viewing time. The collinearity of up-ramp and down-ramp regressors (red crosses), which were used in our analysis, is less than that of up-ramp and boxcar regressors (black crosses). The regressors of participants with longer mean viewing times tended to be somewhat less collinear. For comparison, the dotted lines and solid curves show the collinearity for single unconvolved and convolved regressors, respectively, just as in Fig. 4.

are relatively low (16% and 36% of the noise parameter value of 0.1, whereas 100% or greater is more typical). These results suggest that our low motion coherence and penalty-for-errors manipulation succeeded in boosting DDM decision thresholds to unusually high levels. Such high thresholds yield reasonable accuracy even with low motion coherences, and as a by-product, produce the unusually long response times we sought to induce. Non-decision time estimates are generally expected to be much less than 500 msec (usually in the 200–300 msec range), and therefore constitute a small proportion of the full response times we observed (much smaller than in typical data sets with one-second response times). That the fitting algorithm converged on an estimate of 0 for these parameters is therefore not too surprising.

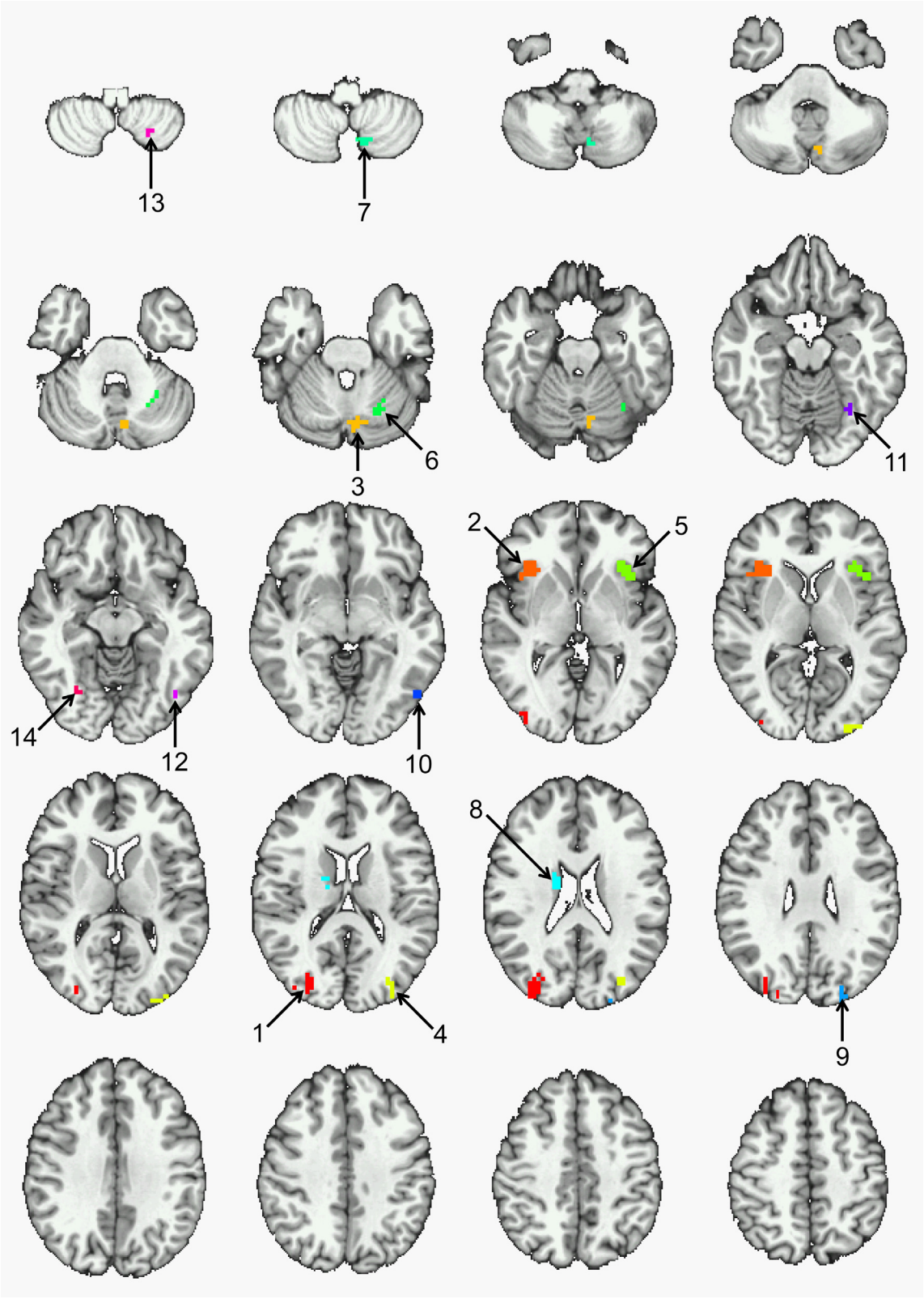

Fig. 8. Areas showing both a main effect of trial type (dots vs. arrows) and an interaction between trial type and ramp regressor type (up-ramp vs. down-ramp) in GLM parameter estimates. Numbers and colors indicate separate clusters, ordered by size, the parameter estimates of which are shown in Fig. 9. (Images are shown in the radiological view, with the left side of each image corresponding to the right side of the head, and with a five-slice gap between subsequent anatomical slices). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. GLM results

Our fMRI analysis used up-ramp and down-ramp regressors, the length of which were determined by the dot motion viewing time for each trial, convolved with the hemodynamic response function. The collinearity of regressors for each participant is shown as red crosses in Fig. 7 (each cross shows the collinearity for one participant). Had we used boxcars instead of down-ramps, the collinearity would have been considerably higher (black crosses). In either case, the collinearity tends to decrease as the mean viewing time for each participant increases. For comparison, the collinearity between these regressors for a single pair of unconvolved regressors (dotted lines) and a single pair of convolved regressors (solid curves) is shown, just as in Fig. 4. The collinearity of regressors for actual participants is lower than the solid curves because there are several regressors for each participant that are jittered with a variable response-to-stimulus interval. Figs. S6–S8 (Supplementary section) compare GLM results for using up-ramp and down-ramp regressors versus up-ramp and boxcar regressors.

We used two criteria to identify areas of the brain involved in evidence accumulation: 1) parameter estimates (beta-weights) must be significantly greater during the perceptual decision making task ("dots" task) than the control task in which participants viewed randomly moving dots and responded to an arrow ("arrows" task); and 2) parameter estimates must be significantly greater for the up-ramp regressor than for the down-ramp regressor, which is indicative of true ramping activity (see Methods). We performed a three-factor repeated measures ANOVA on beta-weights from our GLM analysis, with task type (dots vs. arrows), ramp regressor type (up-ramp vs. down ramp), and participants (modeled as random effects) as our factors (see Methods). Fig. 8 shows areas for which: 1) there was a main effect of task type, and 2) there was an interaction between task type and ramp regressor type (each at a threshold of $p < 0.01$, corrected). Table 2 shows the name, voxel size, and center-of-mass location of all fourteen clusters.

To explore the nature of these interactions, Fig. 9 shows the beta-weights for each cluster, averaged across participants, with error bars indicating the standard error of the mean across participants. In all of these clusters, it turned out that the beta-weights were greater in the dots task than in the arrows task (collapsed across ramp regressor type), and that the difference between beta-weights for the up-ramp regressor and the down-ramp regressor was greater for the dots task than for the arrows task (indicative of greater ramping activity for dots trials than for arrows trials).

We also identified areas of the brain with greater activation during the dots task than the arrows task. These areas, while not necessarily accumulating evidence, are also involved in the decision making process in some way, but could reflect more general processes such as arousal and attention. Fig. S2 and Table S1 (Supplementary section) show areas for which a repeated measures ANOVA revealed larger activation during dots trials than during arrows trials ($p < 0.001$, cluster size ≥ 6). Fig. S3 (Supplementary section) shows the beta-weights for these areas. By definition, in these regions dots trials show greater activation than arrows trials across ramp regressor types, irrespective of ramping. Only a subset of these areas show an interaction between trial type and regressor type. In those areas that do not show a significant interaction, ramping activity is no greater during the decision making task than it is during the control task, and therefore we conclude they are not involved in dot motion evidence accumulation. (Also see Fig. S5 (Supplementary section) for an ROI analysis of area MT, which shows partial overlap with these areas showing greater activity for dots trials than arrows trials. Area MT did not exhibit a significant interaction between condition and ramp type, except when we did not correct for multiple comparisons).

4. Discussion

In this experiment we succeeded in dramatically prolonging the process of perceptual evidence accumulation to the 10-s range relative to the typically observed range of 300 msec–2 s. This was achieved by imposing stiff penalties for incorrect responses. Our analysis of the fMRI data leveraged GLM regressors designed to distinguish ramping activity from boxcar activity. We compared this ramping activity during a perceptual decision making task and during a control task that did not involve evidence accumulation of moving dots but matched the viewing time of dots. Using these methods we identified brain regions involved with perceptual evidence accumulation, which included bilateral anterior insula, cerebellum, caudate, left inferior temporal gyrus, fusiform gyrus, and middle occipital gyrus.

Previous fMRI studies of evidence accumulation during decision making can be divided into several groups. In studies that approach decision making by focusing on areas where activity is modulated by the amount of decision evidence or the type of response effector, the brain regions implicated in evidence accumulation typically include the dorsolateral prefrontal cortex, inferior parietal sulcus, and sometimes the anterior insula (Basten et al., 2010; Domenech and Dreher, 2010, Fillimon et al., 2013, Heekeren et al., 2004, Noppeney et al., 2010; Tosoni et al., 2008). In studies in which decision evidence is revealed progressively over time, reports typically do not include the dorsolateral prefrontal cortex (although see Gluth et al., 2012), but almost always the anterior insula, and sometimes inferior parietal sulcus as well (Gluth et al., 2012; Ivanoff et al., 2008; Ploran et al., 2007, 2011). A potential problem with these previous studies of delaying decisions by gradually revealing a stimulus is that manipulating stimulus evidence is likely to affect a different level of cognitive processing (O'Connell et al., 2012) while confounding perceptual detection with evidence accumulation. Thus, this type of manipulation may not selectively identify the neural integration process (Wheeler et al., 2015). Nevertheless, our observation of accumulator-like activity in the bilateral anterior insula is consistent with these previous studies.

Ploran et al. (2011) attempted to avoid the confound between perceptual detection and evidence accumulation by maintaining a constant amount of perceptual information available. Every two seconds during each trial, a new pixelated mask appeared, which revealed the same area quantity but different parts of the object underneath. Although the visible surface area of the underlying object is constant, the visual information available to the participant is not; the stimulus information is static for two seconds at a time, and then changes discontinuously. This might lead to similarly nonlinear evidence accumulation. This approach may suffice in controlling for the perception/accumulation confound, though it precludes the use of simple, continuous stimuli. Tremel and Wheeler (2015) maintained a constant input of degraded images in a face/house recognition task, but RTs were not sufficiently long (around 3 s) to measure ramping activity. Also, it is possible that face or object recognition is a nonlinear process and that revealing degraded images or masked local parts would interact with the accumulation process in a temporally unstable way. In the future, it may be useful to directly compare these approaches with the methods we propose here of slowing down responding with penalties, to further refine the ability to identify regions selectively associated with evidence accumulation.

The brain regions that are associated with evidence accumulation also depend on the stimulus material used. For example, when decisions involve value judgments, the ventromedial prefrontal cortex is reported (Gluth et al., 2012; Philiastides et al., 2010; Wheeler et al., 2015). In contrast, regions linked to stimulus perception, such as the middle occipital cortex and fusiform gyrus are also reported in various paradigms involving faces, objects, or ran-

Table 2
Areas showing both a main effect of trial type (dots vs. arrows) and an interaction between trial type and ramp regressor type (up-ramp vs. down-ramp).

	Name	Voxels	X	Y	Z
1	R middle occipital gyrus/BA 19	65	−31.3	79.4	17.5
2	R anterior insula	42	−34.4	−20.3	3.7
3	L cerebellum (pyramis)	31	7.3	68.0	−25.0
4	L middle occipital gyrus/BA 19	31	30.0	84.1	13.0
5	L anterior insula	29	33.8	−19.0	4.3
6	L cerebellum (culmen)	19	25.2	56.6	−23.2
7	L cerebellum (inferior semi-lunar lobule)	11	11.6	62.6	−42.2
8	R caudate	10	−15.0	6.6	19.6
9	L middle occipital gyrus/BA 19	10	23.4	85.8	24.7
10	L inferior temporal gyrus/BA37	8	49.1	64.5	−3.5
11	L fusiform gyrus	5	27.9	59.1	−13.1
12	L inferior temporal gyrus/BA37	4	41.2	63.8	−8.0
13	L cerebellar tonsil	3	23.5	56.5	−48.5
14	R fusiform gyrus	3	−27.5	63.5	−9.5

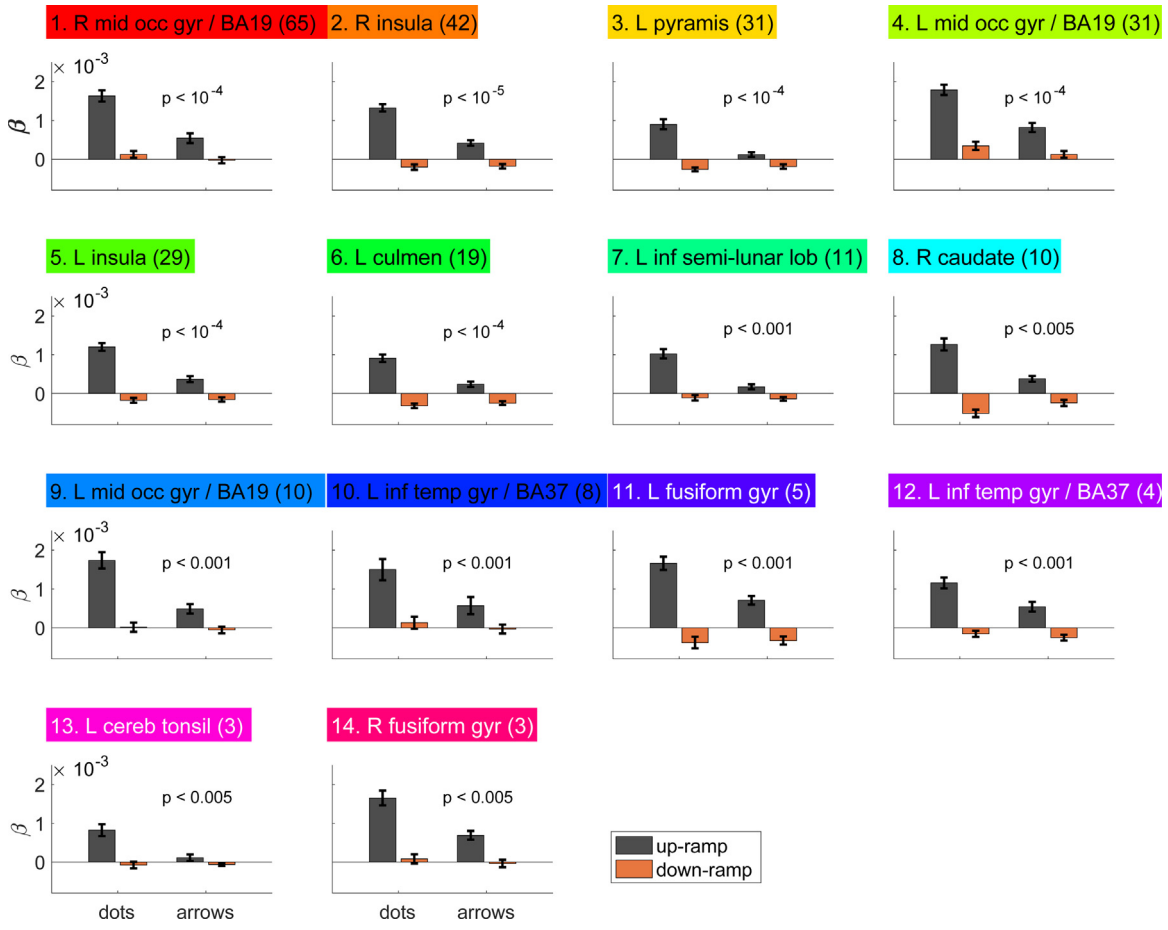

Fig. 9. Parameter estimates (beta-weights) for dots vs. arrows trials and up-ramp vs. down-ramp regressors. Each subplot shows the average beta-weight for a given cluster in Fig. 8, averaged across all participants with error-bars indicating the standard error of the mean. The number in parentheses indicates the number of voxels in a given cluster (cf. Table 2). P-values indicate the Bonferroni-corrected significance of the *t*-test on the interaction between trial type (dots vs. arrows) and regressor type (up-ramp vs. down-ramp).

domly moving dots (Buchsbaum et al., 2013; Ploran et al., 2011; Tosoni et al., 2014; Wheeler et al., 2015). Finally, when decisions are based on memories, areas such as the hippocampus have been shown to be involved (Gluth et al., 2015; Mack and Preston, 2016).

A previous debate in the literature on evidence accumulation during decision making has centered on the question of whether accumulation happens in effector regions, or whether there is effector-independent accumulation. Several studies (e.g., Heekeren et al., 2008; Ploran et al., 2007; Kayser et al., 2010; Meyniel et al., 2013) showed that sensorimotor regions showed activity consistent with a role in evidence accumulation, but also effector-nonspecific areas may be involved (e.g., Ho et al., 2009; Filimon et al., 2013; Liu and Pleskac, 2011; Sestieri et al., 2014). Nevertheless, it may be the case that the distinction between pre-effector and post-effector stages is artificial (van Vugt et al., 2014; Selen et al., 2012), such that decisions are made through a continuous flow of information from regions more involved in perceptual information processing to regions more involved in motor processes. In our study, having successfully prolonged evidence accumulation to a degree that motor implementation takes up only a small part of the stimulus-to-response time, we do not observe accumulation in motor-related areas.

Some have suggested that time on task is a confound (e.g. Buchsbaum et al., 2013; Grinband et al., 2011; Yarkoni et al., 2009). According to this argument, some brain activity may come from the passage of time, and not evidence accumulation per se. To avoid this potential confound, Buchsbaum et al. (2013) used a stop-signal design to dissociate time on task from evidence accumulation, and showed that more posterior perceptual regions were primarily associated with time on task, while more frontal regions such as preSMA and anterior insula were more correlated with the amount of actual evidence that was accumulated. In another design, Noppeney et al. (2010) argued that the inferior frontal sulcus showed characteristics consistent with evidence accumulation during a visual selective attention task with auditory distractors. This is consistent with our findings of accumulator-like activity in the bilateral anterior insula, although we do observe ramping activity also in more perceptual regions such as the middle occipital gyrus and the fusiform gyrus.

Some of the original studies in monkeys using dot motion stimuli found a correlation between ramping activity in the lateral intraparietal cortex (area LIP; e.g. Shadlen and Newsome, 2001) which was interpreted as evidence that this area may drive perceptual decision making. Surprisingly, we failed to find ramping activity in this area in humans. However, a recent study found that pharmacological inactivation of area LIP failed to have any impact on performance in monkeys performing the same task (Katz et al., 2016). Similar reports have been made by others (e.g., Balan and Gottlieb, 2009). These results call into question whether area LIP in monkeys plays a causal role in perceptual decisions. Similarly, activity in human area LIP may correlate with but not play a causal role in evidence accumulation. Our findings are consistent with this interpretation. When we conducted an analysis to identify the main effect of a ramp (contrasting up-ramp against down-ramp activity) irrespective of condition (i.e., collapsing over experimental and control conditions), we identified large areas of the parietal cortex, including area LIP, which were not observed when we contrasted the two conditions. This suggests that ramping in LIP was a general characteristic of prolonged responding, and not directly related to the accumulation of evidence from the dot-motion stimuli in the experimental condition.

It is important to note that our findings do not conclusively establish that participants in our task were accumulating evidence truly continuously. An alternative explanation could involve a sequence of different accumulation processes punctuated by moments when the participant deliberates about whether it is presently worthwhile to respond (Graziano et al., 2011; Ratcliff and Rouder, 1998). While our data were well fitted by a continuous evidence accumulation model, differing variants of integrator models are typically quite difficult to disentangle on a neural level (Ditterich, 2010).

It is also worth noting that the results do not distinguish low coherence trials and high coherence trials. When coherence level is included in the model, there is no significant difference between the two (using the same significance thresholds as in our other results). One possible explanation is that while high coherence trials should have stronger accumulation signals (the BOLD signal rising more rapidly from baseline), the model fits might actually be weaker because high coherence trials entail shorter RTs on average, which implies less area under the curve and greater collinearity between regressors. Conversely, while low coherence trials should involve weaker accumulation signals in the brain, the model might actually fit BOLD accumulation signals better because there is less collinearity between regressors (because the RTs tend to be longer). This combination of stronger brain signals but worse model fits for high coherence trials, and vice versa for low coherence trials, may make it difficult to distinguish the two.

While evidence accumulation is one potential source of ramping activity in perceptual decision tasks, other possible sources could be timing processes, motivational salience, urgency, and/or (potentially associated) increasing punishment-related anxiety. Without being able to distinguish low coherence and high coherence trials, it is difficult to rule these possibilities out. Cisek et al. (2009) suggest that ramping activity from perceptual input is a result of the multiplication of an urgency-to-respond signal with perceptual flows of information. With regard to timing, we have previously suggested that there is a tight link between timing and evidence accumulation (Simen et al., 2011; Balci and Simen, 2014; Simen et al., 2016), which may be amplified for the long time intervals used here. It is worth noting that among the fMRI studies of evidence accumulation, our study is one of a few (see also Ploran et al., 2007) demonstrating cerebellar activity consistent with evidence accumulation. As time progresses, the participant may be more and more inclined to press the button, because they know that in order to gain points, they need to perform as many trials as possible. However, the punishment on errors we introduced made it such that participants are relatively hesitant to commit to a response. In fact, we observed that some participants found the task stressful because of this penalty and the associated anxiety. Activity in the anterior insula has previously been linked to the magnitude and likelihood of impending punishments (Ullsperger et al., 2014) and choice-related anxiety (Shenhav and Buckner, 2014), which is consistent with mounting activation in this region observed during the protracted decision process in our task. The representation of such motivationally significant information in anterior insula is thought to influence the allocation of control signals through the dorsal anterior cingulate cortex (dACC) (Bush et al., 2000; Craig, 2009; Singer et al., 2009; Medford and Critchley, 2010; Ullsperger et al., 2010; Shackman et al., 2011; Shenhav et al., 2013), and could be used by dACC to signal and/or adjust to conflict between responding and response inhibition (Wiecki and Frank, 2013).

In summary, we introduced a new way to dramatically slow down decision making, thereby making evidence accumulation theoretically observable on the time scale of fMRI recordings. Using this method, we observed BOLD activity consistent with evidence accumulation for perceptual decisions in bilateral anterior insula, cerebellum, caudate, left inferior temporal gyrus, fusiform gyrus, and middle occipital gyrus. These findings challenge the interpretation of some areas previously associated with evidence accumulation in perceptual decision making, while identifying some new areas as candidates for involvement in evidence accumulation, integration, and/or closely related processes. Future studies using these methods may help dissociate the specific role that these areas play in perceptual decision making.

Acknowledgements

The authors gratefully acknowledge funding from the AFOSR Multidisciplinary Research Program of the University Research Initiative, FA 9550-07-1-0537 (to Paul M. Krueger, Marieke K. van Vugt, Leigh Nystrom, Philip Holmes and Jonathan D. Cohen) and NIMH grant F32MH080524 (to Patrick Simen). The funding sources were not involved in the study design, data collection, analysis or interpretation, writing the manuscript, or deciding where to submit the work. The authors would also like to thank Peter Foster and Laura deSouza for help with testing participants. The authors have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2017.01.012.

References

Balan, P.F., Gottlieb, J., 2009. Functional significance of nonspatial information in monkey lateral intraparietal area. J. Neurosci. 29 (25), 8166–8176.

Balci, F., Simen, P.A., Niyogi, R., Saxe, A., Hughes, J., Holmes, P., Cohen, J.D., 2011. Acquisition of decision making criteria: reward rate ultimately beats accuracy. Atten. Percept. Psychophys. 73, 640–657.

Basten, U., Biele, G., Heekeren, H.R., Fiebach, C.J., 2010. How the brain integrates costs and benefits during decision making. Proc. Natl. Acad. Sci. 107 (50), 21767ð21772.

Belshey, D.A., Kuh, E., Welsh, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, Inc, New York.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J.D., 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. 113 (4), 700–765.

Bogacz, R., Hu, P.T., Holmes, P., Cohen, J.D., 2010. Do humans produce the speed-accuracy tradeoff that maximizes reward rate? Q. J. Exp. Psychol. 63 (5), 863–891.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 443–446.

Britten, K.H., Shadlen, M.N., Newsome, W.T., Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12 (12), 4745–4765.

Brown, S.D., Heathcote, A., 2008. The simplest complete model of choice reaction time: linear ballistic accumulation. Cognit. Psychol. 57, 153–178.

Buchsbaum, B.R., Erickson, D.T., Kayser, A.S., 2013. Decomposing effects of time on task reveals an anteroposterior gradient of perceptual decision regions. PLoS One 8 (8), e72074.

Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4 (6), 215–222.

Cisek, P., Puskas, G.A., El-Murr, S., 2009. Decisions in changing conditions: the urgency-gating model. J. Neurosci. 29 (37), 11560–11571.

Cox, R.W., Reynolds, R.C., Taylor, P.A., 2016. AFNI and clustering: false positive rates redux. bioRxiv, 065862.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Craig, A.D., 2009. How do you feel—now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10 (1).

Ditterich, J., 2010. A comparison between mechanisms of multi-alternative perceptual decision making: ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. Front. Neurosci. 4, 184.

Donner, T., Siegel, M., Fries, P., Engel, A.K., 2009. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Curr. Biol. 19, 1581–1585.

Filimon, F., Philiastides, M.G., Nelson, J.D., Kloosterman, N.A., Heekeren, H.R., 2013. How embodied is perceptual decision making? evidence for separate processing of perceptual and motor decisions. J. Neurosci. 33 (5), 2121–2136.

Gluth, S., Rieskamp, J., Büchel, C., 2012. Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. J. Neurosci. 32 (31), 10686–10698.

Gluth, S., Sommer, T., Rieskamp, J., Büchel, C., 2015. Effective connectivity between hippocampus and ventromedial prefrontal cortex controls preferential choices from memory. Neuron 86 (4), 1078–1090.

Gold, J.I., Shadlen, M.N., 2000. Representation of a perceptual decision in developing oculomotor commands. Nature 404, 390–394.

Gold, J.I., Shadlen, M.N., 2001. Neural computations that underlie decisions about sensory stimuli. Trends Cogn. Sci. 5 (1), 10–16.

Graziano, M., Polosecki, P., Shalom, D.E., Sigman, M., 2011. Parsing a perceptual decision into a sequence of moments of thought. Front. Integr. Neurosci. 5, 45.

Grinband, J., Savitskaya, J., Wager, T.D., Teichert, T., Ferrera, V.P., Hirsch, J., 2011. The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. Neuroimage 57 (2), 303–311.

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. Proc. Natl. Acad. Sci. 103 (26), 10023–10028.

Heekeren, H.A., Marrett, S., Ungerleider, L.G., 2008. The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci. 9, 467–479.

Ho, T., Brown, S., Serences, J.T., 2009. Domain general mechanisms of perceptual decision making in human cortex. J. Neurosci. 29 (27), 8675–8687.

Ivanoff, J., Branning, P., Marois, R., 2008. fMRI evidence for a dual process account of the speed-accuracy trade-off in decision-making. PLoS One 3 (7), e2635.

Katz, L.N., Yates, J.L., Pillow, J.W., Huk, A.C., 2016. Dissociated functional significance of decision-related activity in the primate dorsal stream. Nature 535 (7611), 285–288.

Kayser, A.S., Buchsbaum, B.R., Erickson, D.T., D'Esposito, M., 2010. The functional anatomy of a perceptual decision in the human brain. J. Neurophysiol. 103, 1179–1194.

Liu, T., Pleskac, T., 2011. Neural correlates of evidence accumulation in a perceptual decision task. J. Neurophysiol. 106, 2383–2398.

Mack, M.L., Preston, A.R., 2016. Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. Neuroimage 127, 144–157.

Medford, N., Critchley, H.D., 2010. Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. Brain Struct. Funct. 214 (5–6), 535–549.

Meyniel, F., Sergent, C., Rigoux, L., Daunizeau, J., Pessiglione, M., 2013. Neurocomputational account of how the human brain decides when to have a break. Proc. Natl. Acad. Sci. 110 (7), 2641–2646.

Mulder, M.J., Van Maanen, L., Forstmann, B.U., 2014. Perceptual decision neurosciences–a model-based review. Neuroscience 277, 872–884.

Noppeney, U., Ostwald, D., Werner, S., 2010. Perceptual decisons formed by accumulation of audiovisual evidence in prefrontal cortex. J. Neurosci. 30 (21), 7434–7446.

O'Connell, R.G., Dockree, P.M., Kelly, S.P., 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15 (12), 1729–1735.

Philiastides, M.G., Biele, G., Heekeren, H.R., 2010. A mechanistic account of value computation in the human brain. Proc. Nat. Acad. Sci. U. S. A. 107 (20), 9430–9435.

Ploran, E.J., Nelson, S.M., Velanova, K., Donaldson, D.I., Petersen, S.E., Wheeler, M.E., 2007. Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. J. Neurosci. 27 (44), 11912–11924.

Ploran, E.J., Tremel, J.J., Nelson, S.M., Wheeler, M.E., 2011. High quality but limited quantity perceptual evidence produces neural accumulation in frontal and parietal cortex. Cereb. Cortex 21, 2650–2662.

Ratcliff, R., Rouder, J., 1998. Modeling response times for two-choice decisions. Psychol. Sci. 9, 347–356.

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108.

Ratcliff, R., 2002. A diffusion model account of response time and accuracy in a brightness discrimination task: fitting real data and failing to fit fake but plausible data. Psychon. Bull. Rev. 9, 278–291.

Selen, L.P.J., Shadlen, M.N., Wolpert, D.M., 2012. Deliberation in the motor system: reflex gains track evolving evidence leading to a decision. J. Neurosci. 32 (7), 2276–2286.

Sestieri, C., Tosoni, A., Mignogna, V., McAvoy, M.P., Shulman, G.L., Corbetta, M., Luca Romani, G., 2014. Memory accumulation mechanisms in human cortex are independent of motor intentions. J. Neurosci. 34 (20), 6993–7006.

Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12 (3), 154–167.

Shadlen, M.N., Newsome, W.T., 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J. Neurophysiol. 86 (4), 1916–1936.

Shenhav, A., Buckner, R.L., 2014. Neural correlates of dueling affective reactions to win–win choices. Proc. Natl. Acad. Sci. 111 (30), 10978–10983.

Shenhav, A., Botvinick, M.M., Cohen, J.D., 2013. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79 (2), 217–240.

Simen, P., Balci, F., Cohen, J.D., Holmes, P., 2011. A model of interval timing by neural integration. J. Neurosci. 31 (25), 9238–9253.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., Cohen, J.D., 2009. Reward-rate optimization in two-alternative decision making: empirical tests of theoretical predictions. J. Exp. Psychol. Hum. Percept. Perform. 35, 1865–1897.

Simen, P., Vlasov, K., Papadakis, S., 2016. Scale (in)variance in a unified diffusion model of decision making and timing. Psychol. Rev. 123, 111–191.

Singer, T., Critchley, H.D., Preuschoff, K., 2009. A common role of insula in feelings, empathy and uncertainty. Trends Cogn. Sci. 13 (8), 334–340.

Starns, J.J., Ratcliff, R., 2010. The effects of aging on the speed- accuracy compromise: boundary optimality in the diffusion model. Psychol. Aging 25, 377–390.

Tosoni, A., Corbetta, M., Calluso, C., Committeri, G., Pezzulo, G., Romani, G.L., Galati, G., 2014. Decision and action planning signals in human posterior parietal cortex during delayed perceptual choices. Eur. J. Neurosci. 39 (8), 1370–1383.

Tosoni, A., Galati, G., Romani, G.L., Corbetta, M., 2008. Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. Nat. Neurosci. 11 (12), 1446–1453.

Ullsperger, M., Harsay, H.A., Wessel, J.R., Ridderinkhof, K.R., 2010. Conscious perception of errors and its relation to the anterior insula. Brain Struct. Funct. 214 (5–6), 629–643.

Ullsperger, M., Danielmeier, C., Jocham, G., 2014. Neurophysiology of performance monitoring and adaptive behavior. Physiol. Rev. 94 (1), 35–79.

VandeKerckhove, J.A., Tuerlinckx, F., 2007. Fitting the Ratcliff diffusion model to experimental data. Psychon. Bull. Rev. 14 (6), 1011–1026.

VandeKerckhove, J.A., Tuerlinckx, F., 2008. Diffusion model analysis with MATLAB: a DMAT primer. Behav. Res. Methods 40 (1), 61–72.

van Vugt, M.K., Simen, P., Nystrom, L., Holmes, P., Cohen, J.D., 2012. EEG oscillations reveal neural correlates of evidence accumulation. Front. Hum. Neurosci. 6 (106), 108–120.

van Vugt, M.K., Simen, P., Nystrom, L., Holmes, P., Cohen, J.D., 2014. Lateralized readiness potentials reveal properties of a neural mechanism for implementing a decision threshold. PLoS One 9 (3), e90943.

Wheeler, M.E., Woo, S.G., Ansel, T., Tremel, J.J., Collier, A.L., Velanova, K., Ploran, E.J., Yang, T., 2015. The strength of gradually accruing probabilistic evidence modulates brain activity during a categorical decision. J. Cogn. Neurosci. 27 (4), 705–719.

Wiecki, T.V., Frank, M.J., 2013. A computational model of inhibitory control in frontal cortex and basal ganglia. Psychol. Rev. 120 (2), 329–355.

Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., Braver, T.S., 2009. BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4 (1), e4257.