MASTER'S THESIS

Neural correlates of subjective confidence in visual decision-making

Laura Geurts
s4419294
l.geurts@donders.ru.nl

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

in
Artificial Intelligence

at
Radboud University
Faculty of Social Sciences

December 1, 2016
Nijmegen, the Netherlands

Supervised by

Dr. Janneke Jehee
Donders Institute for Brain, Cognition and Behavior
Radboud University

Dr. Marcel van Gerven
Donders Institute for Brain, Cognition and Behavior
Radboud University

Abstract

Humans are able to reflect on the quality of their own decisions. With each of our decisions, we associate a certain degree of confidence. Moreover, recent work suggests that human observers use internal estimates of sensory uncertainty when making decisions. This raises the question whether confidence is a readout of such estimates, and if so, how and where confidence is computed from uncertainty. Earlier research has suggested a variety of brain areas potentially involved in the computation of confidence. However, in those studies, stimulus properties were modulated to induce variability in confidence, thereby providing subjects with cues regarding stimulus difficulty and thus expected performance. Here, subjects performed a visual decision task in which stimulus properties were held constant across trials, such that confidence reports could only be based on internal measures. Our data show that even without variability in stimulus difficulty, humans are able to reflect on their own performance, suggesting that confidence is at least partially based on internal estimates of sensory uncertainty. We used fMRI to identify brain regions encoding confidence during this task, and mainly found confidence-related activation in the striatum, particularly in the head of the caudate nucleus and the nucleus accumbens. Our results suggest that this region plays a key role in the computation of confidence from internal estimates of uncertainty. This work provides leads for future research on the neural mechanisms underlying the computation of decision confidence and the role of uncertainty in neural computations in general.

Contents

List of Abbreviations

2AFC	2-alternative forced choice
BOLD	Blood-oxygen-level dependent
(f)MRI	(Functional) magnetic resonance imaging
FOV	Field of view
FWHM	Full width at half maximum
GLM	General linear model
HRF	Hemodynamic response function
LIP	Lateral intraparietal cortex
M1	Primary motor cortex
MoA	Method-of-adjustment
MPRAGE	Magnetization-prepared rapid gradient echo
OFC	Orbitofrontal cortex
PFC	Prefrontal cortex
PMC	Premotor cortex
S1	Primary somatosensory cortex
S2	Secondary somatosensory cortex
TE	Echo time
TMS	Transcranial magnetic stimulation
TR	Repetition time

1 Introduction

1.1 Sensory uncertainty

Every day, we are forced to make many – smaller or larger – decisions. Think, for example, of when you are riding a bike and the traffic light you are approaching switches from green to orange. Now you need to decide whether or not to brake. There will be several factors influencing this decision, such as prior knowledge (i.e. knowledge of traffic rules, previous experiences), your current goals (i.e. being in a hurry or not), but also perceptual information about the current state of your environment, such as your distance to the traffic light and to other traffic. Information about our surroundings is generally *indirect*. Based on sensory input, the brain constructs an internal model of the external world. This reconstruction is not always equally accurate. In the traffic light example, the accuracy of estimating the distance to the intersection might depend on light and weather conditions, but also on your general arousal stater. On top of that, neuronal signal processing by itself is noisy, meaning that some random variability is always introduced into the signal. Hence, all perceptual decisions are subject to uncertainty due to external (stimulus-related, i.e. luminance) as well as internal (neuronal) noise.

How we make decisions given this inherent ambiguity of sensory signals and neuronal computations, is one of the most fundamental, yet unanswered, questions in (computational) neuroscience. Much of the recent work in this field has its foundations in Bayesian probability theory. Bayesian inference refers to a process in which the most likely state of the world is *inferred* from various sources of knowledge and the degree of uncertainty in that knowledge (Vilares & Körding, 2011; Knill & Pouget, 2004; Vilares & Körding, 2011; Pouget, Beck, Ma, & Latham, 2013). Taking into account the degree of uncertainty associated with individual pieces of information is the common feature across Bayesian theories. The following example illustrates why it is worthwhile taking into account this uncertainty: again, you are riding your bike and approaching a traffic light. Now, assume that the traffic light has already turned red, but you are in a rush, since you are running late for an important meeting. However, skipping the light could get you involved in a traffic accident. Thus, you have to estimate the most probable outcome given that you stop or do not stop for the traffic light. To estimate the risk of an accident, you might rely on different sources of information, such as visual information (do I see any cars coming?), auditory information (do I hear any cars coming?), and prior knowledge (do I know this intersection to be a busy one?). Normally, we will rely more on the up-to-date empirical information than on prior knowledge, and visual information tends to be more accurate than auditory information when it comes to localizing objects, as human vision is more spatially accurate than human hearing. Thus, assuming you have a good view on the intersection, then even if you know the road you are about to cross can be busy,

and you hear the sound of engines, as long as you do not see any traffic nearby, you will probably decide to cross. However, if it happens to be foggy or dark, or there are some trees obstructing your view of the intersection, then visual input becomes less informative and thus auditory information and prior knowledge are relatively more reliable. Given the same prior knowledge and auditory input, you might now decide to stop. Hence, when combining different pieces of information in order to make a decision, it is useful to take into account the amount of uncertainty associated with each piece of information and to weigh them accordingly. This is exactly what Bayesian theories argue for. In mathematical terms, Bayesian theories represent knowledge and information as probability distributions, where the mean represents the estimated value of the variable of interest, and the width of the distribution is taken as a metric on uncertainty. Using these probability distributions, Bayesian theories provide a description of how to optimally make decisions. When talking about a Bayesian observer, we think of an ideal observer, who takes into account all available information as efficiently as possible.

Human behavior often resembles Bayes optimal behavior, as has been shown in various studies on, for example, cue combination (e.g. Ernst and Banks, 2002 and sensorimotor learning (e.g. Körding and Wolpert, 2004). Thus, the brain appears to perform Bayesian inference and take into account uncertainty at intermediate steps in processes leading to perceptual decision-making and/or motor actions. How exactly probability distributions are represented at the neural level, remains an open question, but the prevalent idea is that sensory uncertainty can be read out from population activity patterns (Ma, Beck, Latham, & Pouget, 2006; Pouget et al., 2013). Neurophysiological data supporting the idea that uncertainty is encoded in the activity of neuronal populations have been presented for lateral intraparietal cortex (Beck et al., 2008), dorsal medial superior temporal cortex (Fetsch, Pouget, DeAngelis, & Angelaki, 2012), and primary visual cortex (Orbán, Berkes, Fiser, & Lengyel, 2016). Our lab has recently developed a novel method to decode probability distributions from the early visual cortex in humans, using fMRI (van Bergen, Ma, Pratte, & Jehee, 2015). Decoded uncertainty and behavioral errors were found to be correlated on a single-trial basis, which suggests that it is possible to extract information about uncertainty in sensory representations from BOLD activity using this method. Moreover, a comparison between decoded uncertainty and behavioral data implies that subjects adapt their behavior as a consequence of the degree of (decoded) uncertainty in their internal sensory representations. Important to note is that stimulus properties were held constant, i.e. no noise was added to the stimuli themselves, such as by image blurring or contrast modulation. Any variation in stimulus uncertainty as well as behavioral performance may thus be attributed to internal noise (also see section 1.4). Thus, it appears that both humans and animals make use of estimates of uncertainty in decision-making, and this information may be encoded in the activity patterns of neuronal populations.

1.2 Uncertainty and subjective confidence

If the brain indeed accounts for estimated uncertainty in its computations, and sensory uncertainty affects our behavior, then that raises the question whether we have conscious access to measures of uncertainty. Generally, we are able to reflect on the quality or reliability of our own decisions, and associate a certain degree of confidence with each decision we make (Grimaldi, Lau, & Basso, 2015). Before continuing, it is important to point out some differences between confidence and (un)certainty. First of all, confidence only exists in the context of a decision, whereas uncertainty is a property of any probability distribution, irrespective of whether a decision is made (Pouget, Drugowitsch, & Kepecs, 2016). Thus, uncertainty exists at all intermediate stages of probabilistic inference whereas confidence only refers to the probability of correctness of the final choice. Secondly, uncertainty is a property of a distribution and multiple uncertainties may be relevant simultaneously (i.e. for multiple sensory domains or sources of information), whereas confidence is merely a single scalar value associated with the current choice (Pouget et al., 2016). This type of confidence has therefore also been referred to as *summary confidence* as opposed to *distributional confidence* (uncertainty) (Meyniel, Sigman, & Mainen, 2015). Here, we also refer to it as *subjective confidence*, to emphasize the distinction between the actual uncertainty or noise affecting our decisions from our subjective experience of confidence. The question remains how, where, and when confidence is read out, and how exactly it relates to uncertainty, both behaviorally and neurally.

1.3 Functional anatomy of confidence

A wide network of brain structures has been associated with decision-making. Some of these structures are only involved in decisions involving certain sensory modalities or motor components, while others play more a more general role (Gold & Shadlen, 2007). Given that Bayesian brain theories suppose that uncertainty should be taken into account at every computational step in the decision process, we would expect uncertainty to be represented in all of these structures. However, that still leaves the question where and how the degree of confidence about the final decision is computed.

Brain areas typically associated with perceptual confidence are mostly located in the prefrontal cortex (PFC). In one of the first studies focusing explicitly on decision confidence, rats were trained on an odor mixture categorization task (Kepecs, Uchida, Zariwala, & Mainen, 2008). Firing rates of neurons in the orbitofrontal cortex (OFC) were found to correlate with stimulus uncertainty, and given the same amount of uncertainty, their average firing rates were higher on correct trials than on incorrect ones. Moreover, the rat's willingness to wait for a delayed reward was reverse correlated with stimulus uncertainty, indicating that with lower uncertainty, decision confidence increased. Also, pharmacological inactivation of OFC has been found to impair willingness to wait (decision confidence), but not choice accuracy (Lak et al., 2014). Altogether, these observations suggest that in rats, OFC activity represents a measure of confidence. In humans,

PFC has also been associated with uncertainty and confidence in decision-making (De Martino, Fleming, Garrett, & Dolan, 2013). In a value-based decision-making task, subjects were asked to choose between food items to consume later, and to report confidence for each decision. Both ventromedial and rostrolateral PFC showed confidence-related activity, and the authors conclude that ventromedial PFC appears to be responsible for the qualitative evaluation of decisions, feeding into rostrolateral PFC, which reads out confidence from this activity. This fits well with the rat studies, since in humans, ventromedial PFC includes the OFC, and moreover human ventromedial PFC has been suggested as the functional analogue to rat OFC. This is not the only study suggesting a role for the PFC in confidence evaluation. Individual differences in metacognitive ability – i.e. the correspondence between objective accuracy and subjective confidence on a simple task – were found to correlate with gray matter density and white matter microstructure in anterior PFC (Fleming, Weil, Nagy, Dolan, & Rees, 2010). Moreover, patients with lesions in anterior PFC, when compared to healthy controls, showed impaired metacognitive ability on a visual task, but normal objective accuracy (Fleming, Ryu, Golfinos, & Blackmon, 2014). Thus, in humans as well as animals, several areas within the PFC have been functionally associated with metacognitive evaluation and confidence.

However, prefrontal cortex is not the only area that has been suggested to encode confidence. Using electrophysiological techniques in monkeys, confidence-related activity has been shown in several other areas. For example, lateral intraparietal cortex (LIP) has been associated with confidence in a visual decision task (Kiani & Shadlen, 2009). In this task, monkeys made decisions about the motion direction of moving dot stimuli and were rewarded for correct decisions. Confidence was measured behaviorally by offering an 'opt-out' option on some trials: the animal was then offered to either make a motion direction decision, or to receive a smaller but certain reward. Opting out was interpreted as a sign of low decision confidence. LIP spiking activity correlated with both the direction of decision and confidence. In a very similar task, neurons in the pulvinar nucleus of the thalamus also responded to confidence, but not to the decision direction (Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013). Accordingly, pharmacological inactivation of the pulvinar affected confidence judgments but not decision accuracy. Another area that has been found to encode confidence in its spiking activity is the supplementary eye field (Middlebrooks & Sommer, 2012). In this case, confidence was measured by having the animals bet on the correctness of each decision. Potential neural correlates of confidence are thus not restricted to the PFC.

Note, however, that all of these studies modulate stimulus difficulty in order to trigger variation in confidence. This introduces possible confounds which I will elaborate on in the next section. Only recently, a study has been published in which cortical representations of confidence have been investigated while maintaining constant stimulus difficulty across trials (Hebart, Schriever, Donner, & Haynes, 2016). BOLD correlates of subjective confidence in a dot motion decision task were observed in the ventral striatum around the nucleus accumbens, an area classically associated with reward-related processing. The authors speculate that this activity might be related to the rewarding feelings we experience when we are being confident, but this relationship requires further investigation. From the above review of

previous work on confidence and its neural representation, we may conclude that a wide variety of areas appears to play some role in the computation of perceptual confidence. The prefrontal cortex and OFC in particular, have perhaps received most attention in this respect, but correlates of confidence have been seen in areas such LIP, supplementary eye field, the pulvinar, and the striatum, as well.

1.4 Challenges in confidence research

Despite a substantial amount of work having been done, as of yet, there is no real consensus with respect to where in the brain or how decision confidence is read out. A significant limitation of most of the earlier work, as mentioned before, is that stimulus difficulty was usually varied across trials in order to trigger variability in uncertainty and confidence. Although it is obvious that we need across-trial variation in subjective confidence in order to be able to investigate its neural representation, varying stimulus difficulty introduces behavioral confounds. Namely, if one or more physical properties of the stimulus – such as contrast, blur, motion direction coherence, or stimulus duration – are modulated, then subjects might use this as an external cue for the reliability of the sensory information and thus as an indicator of their performance. In other words, perhaps subjects simply learn that high levels of blur typically result in poor behavioral performance, and will therefore report lower confidence for more blurred stimuli. In this scenario, they need not have an explicit representation of perceptual uncertainty – after all, it was the blur that led them to change their decisions. Confidence estimation then reduces to monitoring of external cues, rather than reading out internal uncertainty (Barthelmé & Mamassian, 2010). A paradigm where stimulus difficulty is varied, is thus not suitable if we want to dissociate between the representation of the estimated reliability of the perceptual information (internal uncertainty), and the representation of stimulus properties that subjects associate with their performance on the task (external uncertainty). Therefore, in order to investigate the former, stimulus properties ought to remain constant across trials.

Another issue with many confidence studies is that in most designs – including the work by Hebart et al. (2016) – confidence judgments are only made after the decision itself. Intuitively, this might seem to be the most straight-forward procedure: making a decision and then reflecting upon it. However, if we are interested in the perceptual confidence associated with the actual choice , this approach is somewhat problematic. The reason is that evidence accumulation seems to continue post-decisionally (Pleskac & Busemeyer, 2010), confidence keeps changing over time (Yu, Pleskac, & Zeigenfuse, 2015), and error-monitoring processes may affect postdecisional confidence assessment (Yeung & Summerfield, 2012) and even depend at least partially on the same neural mechanisms (Boldt & Yeung, 2015). Although the the postdecisional development of confidence and the relationship between confidence and error monitoring are interesting topics of research, when we are interested in studying the perceptual confidence at the time of decison-making, we do not want such processes to interfere with our behavioral measures. Hence, postdecisional confidence ratings are

a suboptimal measure of perceptual confidence, and we argue for a design in which confidence reports and decision-making occur simultaneously.

1.5 Project goals

The main goal of the current project was to find BOLD correlates of perceptual confidence as a readout of internal uncertainty. Maintaining constant stimulus difficulty across trials allows us to assume that participants base their confidence judgments on some internal measure of perceptual uncertainty rather than on physical properties of the external stimulus. Moreover, we asked subjects to make a decision and report confidence at the same time, in order to have the confidence judgment linked to the perceptual decision as closely as possible, and prevent postdecisional stimulus processing and error monitoring from interfering with confidence judgments. Compared to previous work, our design is thus more suitable for answering the question where in the brain subjective confidence as a readout of internal sensory uncertainty is encoded.

2 Methods

2.1 Data collection

2.1.1 Participants

23 healthy volunteers (aged 20–30, 11 female) participated in this study and completed the experiment. All participants had normal or corrected-to-normal vision, and provided informed written consent. Seven additional subjects participated but did not complete the experiment: four were excluded after the initial behavioral session due to poor metacognition[1] (3), poor task performance[2] (1) or participation in a TMS experiment just prior to the scan session (1), two started but did not complete the scan sessions due to scanning discomfort (1) or no-show (1).

2.1.2 Task design and stimuli

Participants performed two tasks: a method-of-adjustment (MoA) task and a two-alternative forced choice (2AFC) confidence task. The MoA task is not relevant for any of the questions addressed in this thesis and is therefore not further discussed. For a description of the MoA task, the reader is referred to an earlier publication by our lab (van Bergen et al., 2015).

The design of the 2AFC confidence task is visualized in fig. 2.1. Each task run contained 20 trials (16.5 s each) and two fixation periods at the start (30 s) and the end of the run (60 s). Every trial started with the presentation of the stimulus, a sinusoidal grating (1.5 s), followed by a fixation period of 6 s, after which a reference line was shown for 4.5 s and participants were asked to judge whether the stimulus was rotated clockwise (CW) or counter-clockwise (CCW) with respect to the reference. At the end of each trial, there was another 4.5 s fixation period before the next trial would start. Participants responded using two MR-compatible button boxes, one for the left hand (CCW) and one for the right hand (CW), and were instructed to rest their fingers on the buttons throughout the experiment. Responses were given on a graded scale (1–4) to indicate the level of confidence about the decision: the inner buttons (index fingers) corresponded to low confidence (confidence level 1), while the outer buttons corresponded to high confidence. The stimulus and reference line were identical to the ones used in the preceding study by van Bergen et al. (2015). However, only oblique orientations (45° and 135°) were used here. The reason to exclusively use these orientations, is the fact that orientation perception is known to be more accurate for cardinal than for oblique orientations

[1] Poor metacognition was quantified as a an area under the curve (accuracy versus confidence) < 0.6 in the 2AFC confidence task (see section 2.1.2 for a task description).

[2] Poor task performance was quantified as a standard deviation > 10 degrees on the error distribution of the MoA task (see section 2.1.2 for a task description).

(oblique effect; Appelle, 1972). Thus, in order to maintain constant difficulty across trials, we used the oblique orientations only. Gaussian jitter was added to the stimulus orientations to ensure that the subjects could not use some internal representation of the oblique orientations instead of the actual stimulus. The absolute angular displacement of the reference bar with respect to the grating stimulus was exactly the same in each trial, only varying in terms of direction: CW or CCW. The exact difference between the two was determined on a subject-by-subject basis, as the orientation offset at 75% accuracy in a separate behavioral session. By using only oblique orientations and fixing the angular displacement between the stimulus and the reference, we were able to maintain equal difficulty across trials, such that subjects could not match their confidence judgments to the objective (external) difficulty of the trial. This distinguishes our design from previous studies on confidence and uncertainty and allows us to investigate the influence of internal rather than external uncertainty on confidence.

Stimuli were generated on a Macbook Pro computer using MATLAB and The Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007). During the scan sessions, they were displayed on a rear-projection screen, by a luminance-calibrated EIKI projector (screen resolution: 1024 × 768 pixels, refresh rate: 60 Hz), which was viewed by the subjects through a mirror mounted on the head coil.

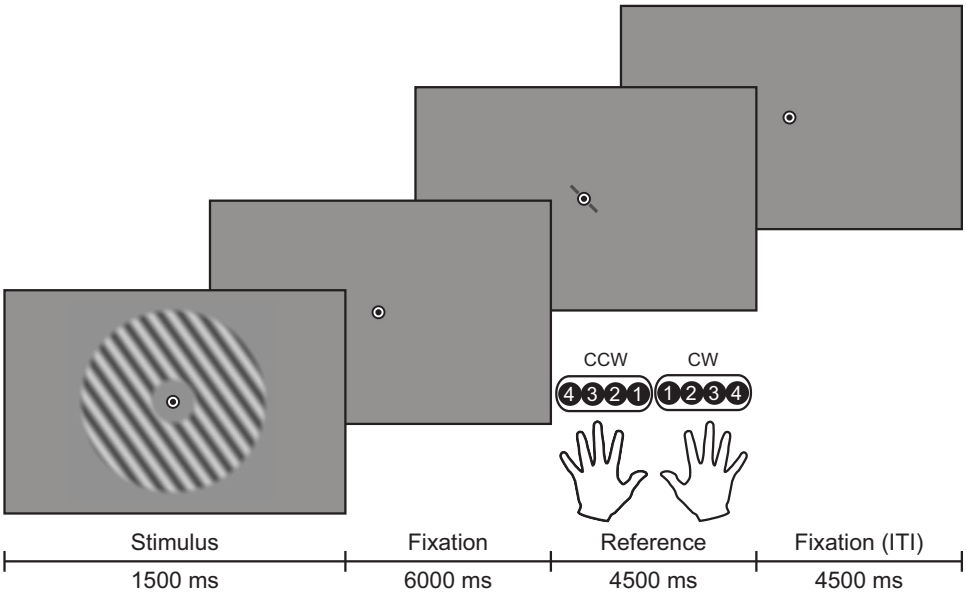

FIGURE 2.1: **Overview of a trial in the 2AFC task.** Each trial starts with the presentation of a stimulus (sinusoidal grating), followed by the presentation of a reference line. Subjects are asked to judge whether the stimulus was rotated clockwise (CW) or counter-clockwise (CCW) with respect to the reference, while indicating their confidence on a 1–4 scale.

2.1.3 Experimental procedures

The study consisted of one behavioral and two scan sessions for each participant. During all sessions, subjects performed both the MoA task and the 2AFC confidence task. The two tasks were alternated in blocks consisting of 20 trials of either one or the other task. Across the two scan sessions, subjects completed a total of 9–12 runs of each task (180–240 trials). Eye movements were tracked and monitored online throughout the scan sessions. The behavioral session, which lasted about 60 minutes, was used to train the participant on the tasks and as an assessment of individual task performance for purposes of establishing individual task difficulty and to verify that subjects' matched our inclusion criteria as described in section 2.1.1.

2.1.4 MRI protocols

MRI data were collected using a Siemens 3T Magnetom Trio scanner with an eight-channel occipital coil, at the Donders Center for Cognitive Neuroimaging (Nijmegen, the Netherlands). At the start of each session, an anatomical image was acquired using a high-resolution MPRAGE protocol (FOV 256 $\times$ 256 mm, 1-mm isotropic voxels). Functional imaging data consisted of 68 transversal slices covering the whole brain, acquired using a T2*-weighted gradient-echo echoplanar imaging protocol (TR 1500 ms, TE 38.60 ms, FOV 210 $\times$ 210 mm, 2-mm isotropic voxels).

2.2 Data analysis

2.2.1 Preprocessing of fMRI data

The raw functional imaging data were motion-corrected with respect to the middle volume of the middle run of the session, using FSL's MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002). Slow drifts in the BOLD signal were removed using a high-pass temporal filter with a cut-off period of 50 s. Images were spatially smoothed with a 6-mm FWHM Gaussian kernel.

Registration from functional to standard space was done in four steps. Functional data were (1) unwarped using B0 fieldmaps to account for magnetic field inhomogeneities and (2) linearly registered to an anatomical image acquired during the same session. Subsequently, (3) this anatomical image was linearly registered to a subject-specific anatomical template created by aligning and averaging the anatomical images from the two separate sessions per subject. Lastly, (4) the subject-specific anatomical template was non-linearly registered to MNI152 standard space. These four transformations were combined to compute the mapping of the functional data to MNI space. FSL's FLIRT was used for the linear registrations (6 degrees of freedom), and FNIRT was used for the non-linear registration (linear prealignment with 12 degrees of freedom, warp resolution: 10 mm) (FLIRT: Jenkinson and Smith, 2001; Jenkinson et al., 2002; Greve and Fischl, 2009, FNIRT: Andersson, Jenkinson, and Smith, 2007).

2.2.2 GLM analyses

The functional data were analyzed with a *general linear model* (GLM), using FSL's FEAT. A GLM is a linear model of the form

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \epsilon$$

Where $\boldsymbol{Y}$ refers to the actual time course of the BOLD response of a single voxel, $\boldsymbol{X}$ is the design matrix containing predictor time courses as column vectors: each column represents a single regressor (independent variable) such as stimulus type or confidence level as it changes over time, $\boldsymbol{\beta}$ is a vector of the weights associated with each of the regressors, and $\boldsymbol{\epsilon}$ is the residual error at each time point: the part of the response ($\boldsymbol{Y}$) which is not explained by the model ($\boldsymbol{X\beta}$). The aim is to find the best possible model fit, i.e. a set of $\boldsymbol{\beta}$ estimates which leaves the smallest possible residual variance $\boldsymbol{\epsilon}$ given the data and the design matrix. These $\boldsymbol{\beta}$s are then our best estimate of the true model parameters. A large β_n (weight) indicates that the BOLD time course of the voxel under investigation is strongly correlated with the state of the nth regressor. For example, if the model includes a regressor indicating all time-points at which a visual stimulus was presented, the weight of this regressor in the model (its β) will be high for voxels in the visual cortex.

Run-level analysis

At the first level of analysis, GLMs were fit per run. Stimulus, response and confidence level where modeled using six regressors – further detailed in table 2.1. Each of the regressors of interest was convolved with a canonical hemodynamic response function (HRF), comprised of two gamma functions – a standard, positive one and a small, delayed, and inverted one, which is used to model the undershoot that follows the initial BOLD increase (Friston et al., 1998). This is a standard approach for modeling the expected event-related BOLD signal. In addition, the first temporal derivatives of each of the six regressors were added to the model. Hemodynamic responses are known to vary across subjects and brain areas, and adding the temporal derivative to the model has been found to partially account for differences in the shape and timing of the BOLD response and thereby improve the fit of the regressors of interest (Handwerker, Ollinger, & D'Esposito, 2004). Lastly, we added an intercept and 24 head motion regressors, together representing the raw displacement parameters as well as their squares and temporal derivatives.

The regressors of interest modeled the stimulus as well as the response within each trial, thereby distinguishing between left-hand and right-hand responses because of the expected differences in their motor components. Moreover, additional regressors were added to model these events for high-confidence trials only, allowing us to investigate the differences in BOLD response between high-confidence and low-confidence trials. The cut-off value for classifying confidence ratings as high or low, was determined for each scan session individually. The cut-off was chosen such that both classes had roughly the same number of trials, i.e. that the difference between the two categories in terms of the number of trials per category was

TABLE 2.1: **List of regressors of interest.** In the final model, each of these regressors was convolved with a canonical HRF.

Regressor	Description
Stimulus_All	Represents the time window of grating stimulus presentation, i.e. the first 1.5 seconds of each trial. This regressor is a box-car function which equals one whenever a stimulus is shown on the screen, and zero otherwise.
Stimulus_HighConf	Same as above, but only for high-confidence trials. This box-car function takes the value of one when a stimulus is presented, but only in those trials where the subject indicated high confidence.
RespLeft_All	Represents left-hand responses (*counter-clockwise*), and is temporally linked to the response time. The regressor equals one only at the time-point when the subject presses the response button.
RespLeft_HighConf	Same as above, but only for high-confidence trials. This function takes on the value of one when the subject responds with the left hand and indicates high confidence.
RespRight_All	Represents right-hand responses (*clockwise*), and is temporally linked to the response time. The regressor equals one only at the time-point when the subject presses the response button.
RespRight_HighConf	Same as above, but only for high-confidence trials. This function takes on the value of one when the subject responds with the right hand and indicates high confidence.

as small as possible (median split). Despite the attempt to keep the number of high- and low-confidence trials roughly equal, there were a few single runs in which not all four response combinations (Left/Right hand–High/Low confidence) occurred. When that happened, two regressors would be perfectly correlated (i.e. when there were no left hand-low confidence responses, the regressors *RespLeft_All* and *RespLeft_HighConf* were identical), which caused problems in model-fitting. Thus, such runs were excluded from further analysis. A total number of 7 runs (in 6 sessions) were excluded for this reason.

Group-level analysis

Contrast maps per run were registered to standard space (MNI) using the transformations mentioned in section 2.2.1. Group-level t-tests were computed in two steps: individual runs were combined on a subject-by-subject basis using a fixed-effects model with one regressor for each session, resulting in one contrast map per subject for each contrast. These maps were subsequently combined across all subjects using a mixed-effects model (FSL's FLAME 1) with a single regressor. The resulting group-level contrast maps were corrected for multiple comparisons using cluster-size thresholding at $p < 0.05$ with a cluster-defining threshold of $p < 0.0001$ (uncorrected).

3 Results

3.1 Behavior

3.1.1 Objective accuracy

Our subjects performed a perceptual decision task in which they were first shown a grating (the stimulus) and were asked to compare this to a subsequently presented bar (the reference) and judge whether the stimulus had been rotated clockwise or counterclockwise with respect to the reference. In order to maintain approximately equal task difficulty across participants, we used an adaptive staircase procedure to ensure a performance level of 75% during the behavioral session. Throughout the scan sessions, we then fixed task difficulty – i.e. the size of offset between the stimulus and the reference – at this level. Indeed, on average individual subjects were correct on 75.6% (S.D. 8.32%) of all trials.

One of the key features of our design is that, although the stimuli were not exactly identical, the objective difficulty of each trial was the same, i.e. no noise was added to the stimuli to induce variability in performance and uncertainty. Our data confirm this assumption. Across trials there was no significant difference in accuracy due to stimulus properties, such as the direction of offset between the stimulus and the reference line (CW/CCW), $t(45) = 0.50$, $p = 0.67$ (fig. 3.1a), or the base orientation of the stimulus (45/135 degrees), $t(45) = -0.43$, $p = 0.67$ (fig. 3.1b). True stimulus orientations were made up of either of these base orientations plus some Gaussian jitter, such that the actual orientations were not exactly equal to 45 or 135 degrees. In order to test whether this jitter affected performance, we binned the trials into three groups based on the absolute angular displacement between the actual stimulus and the base orientation (bin width equaled one-third of the maximum deviation from the base orientation in that particular session). Accuracy did not depend on this deviation from the base, $F(2, 90) = 1.01$, $p = 0.369$ (fig. 3.1c). Altogether, our data support the idea that, as we intended, objective stimulus difficulty was constant across trials in our task.

3.1.2 Confidence

Since we were interested to see whether people have access to information regarding the uncertainty associated with their perceptual decisions, and how this information is represented neurally, we did not only ask subjects to make decisions, but also to rate their their confidence on a 1–4 scale. However, the distribution of responses across this scale varied substantially between individuals: some mostly used the extreme values (1 & 4), whereas others preferred the middle levels, in some cases the distribution was skewed towards one of the extremes, and a few participants only

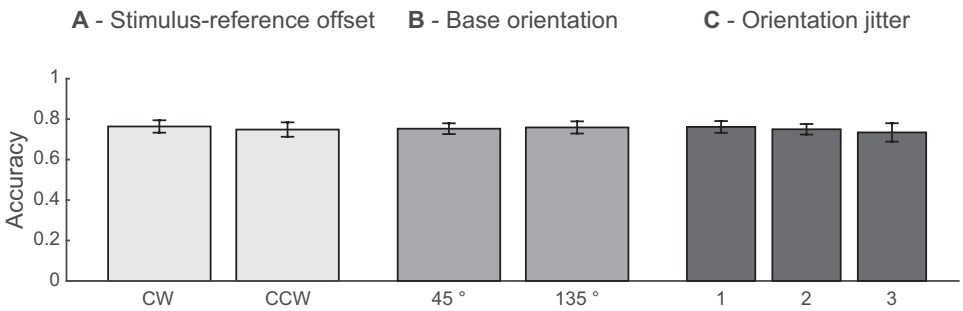

FIGURE 3.1: **Accuracy as a function of stimulus properties.** Individual stimuli differed from each other in three respects: a) the offset of the stimulus with respect to the reference line was either clockwise (CW) or counterclockwise (CCW), b) stimuli were centered around on or the other the oblique axis (45/135 degrees), and c) a small amount of Gaussian jitter was added, such that the actual stimulus orientations slightly deviated from the base orientations. Trials were binned based on the absolute distance between the true orientation and the base orientation of the stimulus, where bin width equaled one-third of the maximum distance in the same session (1 = smallest). Error bars: 95 %-confidence intervals.

used three out of four confidence levels. Thus, in order to correct for inter-individual differences in confidence distributions due to different use or interpretation of the four-point confidence scale, we divided the sample into high- and low-confidence trials. The cut-off value for high versus low confidence was determined for each session separately, using a median split criterion: the cut-off was always chosen such that the within-session difference in frequency between high- and low-confidence trials was as low as possible.

Using this criterion, there was indeed no significant difference between high- and low-confidence trials in terms of frequency, $p = 0.36$, $t(45) = -0.29$. Accuracy was significantly higher on high-confidence ($M = 85.6\%$, $S.D. = 9.4\%$) than on low-confidence trials ($M = 64.9\%$, $S.D. = 10.3\%$), $t(45) = -14.5$, $p < 0.001$ (fig. 3.2a), suggesting that confidence reports were actually meaningful, and that subjects' experienced confidence was indeed predictive of their objective performance. Another well-established behavioral hallmark of confidence in perceptual judgments is that the degree of confidence correlates negatively with reaction times (Volkmann, 1934). This effect was replicated in our dataset (low confidence: $M = 1.91$ s, $S.D. = 0.418$ s, high confidence: $M = 1.55$ s, $S.D. = 0.321$ s), $t(45) = 9.30$, $p < 0.01$ (fig. 3.2b). Thus, despite the lack of variability in objective difficulty across trials, subjects appeared to be able to reflect on their own performance quite accurately.

3.2 Neural correlates of confidence

Our main research question was where in the brain perceptual confidence – as a consequence of internal uncertainty – is read out. Thus we compared BOLD activity during high-confidence trials versus low-confidence trials.

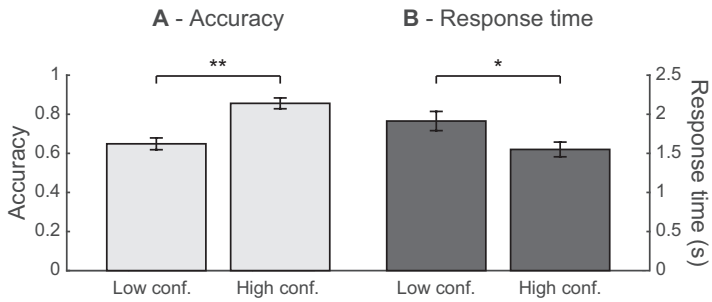

FIGURE 3.2: **Comparing low- versus high-confidence trials** in terms of accuracy (a) and response time in seconds (b). Both accuracy and response times depended on confidence (* $= p < 0.01$, ** $= p < 0.001$). Error bars: 95 %-confidence intervals.

To this end, we conducted a GLM analysis (see section 2.2.2 and table 2.1 for further details regarding the model). We looked at stimulus-related and response-related activity separately. Unless stated otherwise, results are reported at a cluster-wise threshold of $p < 0.05$, family-wise, and a cluster-forming threshold of $p < 0.0001$, uncorrected.

We were primarily interested in fluctuations in subjective confidence due to variability in internal noise. However, visual attention is known to modulate activity in early visual cortex, with an increase in overall BOLD signal in this area when attention is high (Ress, Backus, & Heeger, 2000). Moreover, attention-related activity also predicts performance (Ress et al., 2000). In order to verify that, in our task, between-trial differences in confidence were not merely due to variability in stimulus processing as a consequence of attention, we first compared stimulus-related activity on high- versus low-confidence trials. There were no significant positive or negative differences in BOLD responses to the stimulus depending on confidence, even with a more lenient cluster-forming threshold of $p < 0.001$. This suggests that the variability in reported confidence cannot simply be reduced to fluctuations in visual attention.

We assumed that the readout of perceptual confidence would be temporally linked to the behavioral response (button press). Thus, we looked at the effects of confidence on response-related activity. We did not find any negative correlations of confidence with BOLD activity, but there were several areas where the BOLD signal correlated positively with confidence (table 3.1 and fig. 3.3). First, the striatum showed bilateral activation, primarily around the head of the caudate nucleus and the nucleus accumbens, with some extension into the putamen on the left side only (fig. 3.3a). Furthermore, BOLD signal increased with confidence in the secondary somatosensory cortex (S2), bilaterally, and in the left primary somatosensory cortex (S1), as well as primary motor cortex (M1) and premotor cortex (PMC) (fig. 3.3b). Note that although the activation of S2 was bilateral, the cluster in the left hemisphere was nearly four times the size of the one on the right. Confidence-related activity thus seems to be more pronounced in the left than in the right hemisphere. However, it is worth mentioning here that in our dataset, the relative amount of right-hand responses (as a fraction of the total number of responses), was significantly larger on

high-confidence ($M = 54.5\%$, $S.D. = 12.3\%$) than on low-confidence trials ($M = 47.5\%$, $S.D. = 9.87\%$), $t(45) = -2.98$, $p < 0.01$. Moreover, with a lower cluster-forming threshold of $p < 0.001$, we also found activation of the right M1, S1, and PMC, suggesting that the activation of these areas is in fact bilateral, but appears somewhat left-lateralized due to the relatively higher amount of right-hand responses for high confidence. With this more lenient cluster-forming threshold ($p < 0.001$), we also found some clusters of activation in the supplementary motor area and the insular cortex, but neither of them survived at $p < 0.0001$.

TABLE 3.1: **List of all brain regions in which BOLD signal correlated positively with confidence**, including standard space (MNI) coordinates of all foci separated by at least 10 mm and cluster-wise p-values. Reported are all clusters which survived thresholding at $p < 0.0001$ uncorrected (cluster-forming threshold) and subsequent cluster-size based correction $p < 0.05$ family-wise.

		Peak voxel				Cluster	
L/R	Area	X	Y	Z	Z-val.	Size	P-val.
Left	Secondary somatosensory cortex	-54	-24	20	5.08	192	0.000
Left	Secondary somatosensory cortex	-56	-24	30	4.53		
Left	Secondary somatosensory cortex	-66	-26	26	3.99		
Left	Caudate nucleus (head)/n.accumbens	-8	16	2	4.93	152	0.000
Left	Putamen	-24	14	2	4.29		
Right	Caudate nucleus (head)/n.accumbens	12	18	0	5.10	92	0.003
Left	Primary motor/somatosensory cortex	-40	-24	58	4.77	76	0.005
Left	Premotor cortex	-34	-16	62	4.24		
Right	Secondary somatosensory cortex	66	-18	18	4.88	53	0.017
Right	Secondary somatosensory cortex	54	-18	18	4.48		

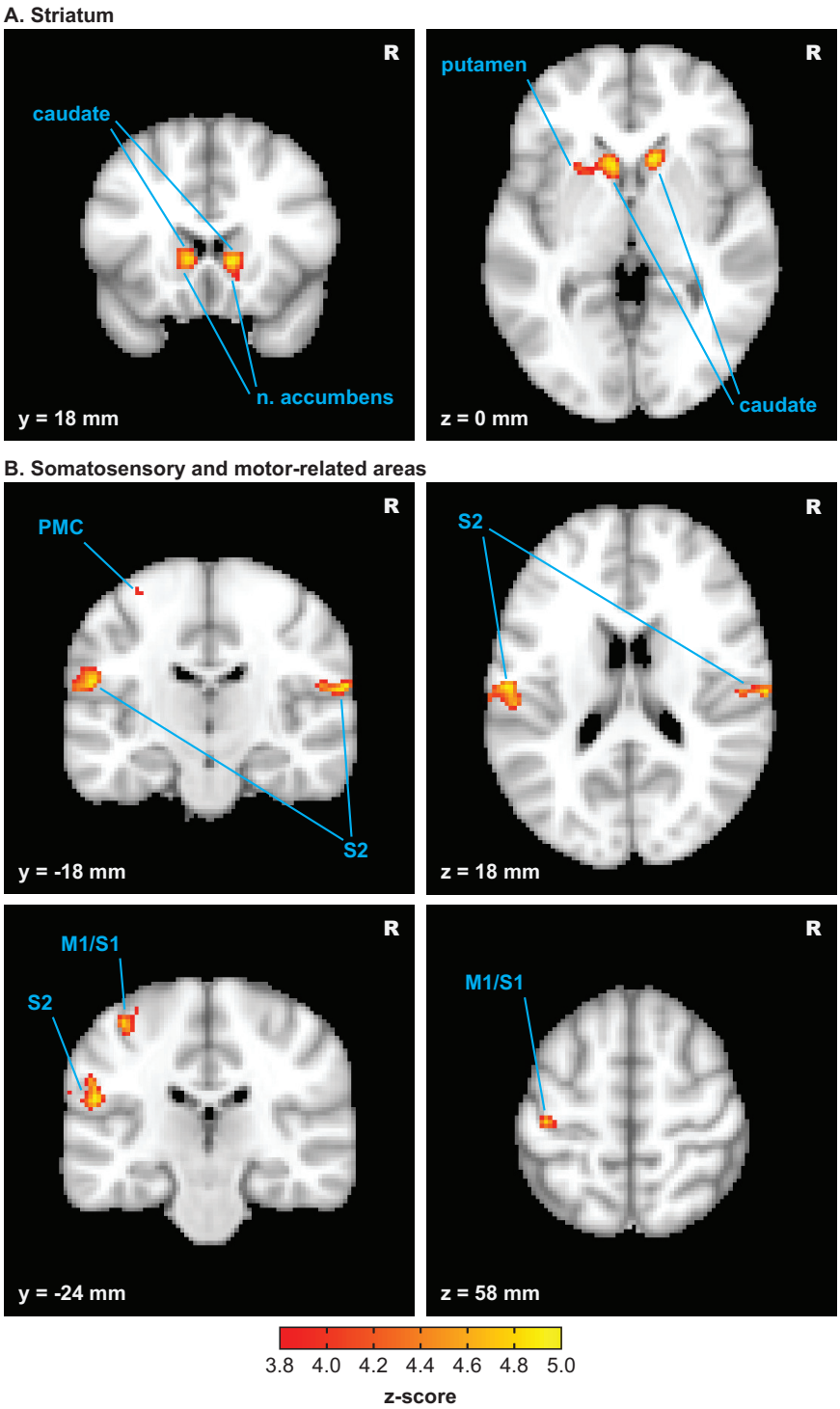

FIGURE 3.3: **Areas showing greater BOLD activity during high-confidence compared to low-confidence decisions.** These regions can be subdivided into: (a) striatum – (head of the) caudate nucleus, nucleus accumbens, and putamen, and (b, next page) sensorimotor areas – secondary somatosensory cortex (S2), primary somatosensory cortex (S1), primary motor cortex (M1) premotor cortex (PMC). All images are thresholded at $p < 0.05$ (cluster-wise, family-wise error corrected), with a cluster-forming threshold of $p < 0.0001$ (uncorrected).

4 Discussion

4.1 Measuring confidence as a readout of uncertainty

There is a growing amount of empirical evidence supporting the idea that in (perceptual) decision-making, the brain's computations are congruent with Bayesian inference This means that in combining different pieces of information, each of those pieces appears to be weighted according to the amount of uncertainty associated with it. Behavioral studies in humans (e.g. Ernst and Banks, 2002; Körding and Wolpert, 2004) as well as neurophysiological work in monkeys (Beck et al., 2008; Fetsch et al., 2012) suggest that both behavior and neural activity associated with perceptual decisions follow the predictions of Bayesian brain theories. On top of this, our lab recently developed a method to decode sensory uncertainty on a single-trial basis from BOLD activity in early visual cortex in humans, and found that observers indeed appear to take this uncertainty into account in their decisions (van Bergen et al., 2015). Thus, our brain seems to not only read out an estimate of the most likely stimulus from the sensory input it receives, but also the degree of uncertainty associated with this estimate. One of the questions these findings raise, is whether we have conscious access to the information about this uncertainty, and if so, how and where uncertainty is read out.

A large amount of work has been done on the subjective confidence associated with (perceptual) decisions and its neural representation, but the evidence remains inconclusive. Various brain regions have been linked to confidence, including several prefrontal areas (Kepecs et al., 2008; Lak et al., 2014; De Martino et al., 2013; Fleming et al., 2010; Fleming et al., 2014), lateral intraparietal cortex (Kiani & Shadlen, 2009), supplementary eye field (Middlebrooks & Sommer, 2012), the pulvinar thalamic nucleus (Komura et al., 2013), and the ventral striatum (Hebart et al., 2016). However, the problem with nearly all of these studies (except for Hebart et al., 2016) is that stimulus difficulty was varied across trials by manipulating stimulus properties, in order to trigger variability in perceptual confidence. The problem with such a design is that, rather than reading out the uncertainty from the sensory signal, subjects might learn to link properties of the external stimulus to their performance. For example, if on a particular trial the stimulus has relatively low contrast and the subject is aware of this, then they might realize that their chance of making a correct decision for this stimulus will be lower and thus report lower confidence. In this case, it is impossible to judge whether the measured confidence really is a readout of internal uncertainty or, instead, simple stimulus monitoring. In order to address the question whether we are aware of the internal measure of uncertainty, we should thus keep stimulus properties constant in order to be sure that subjective confidence judgments are not simply based on monitoring of physical features of the external stimulus. Moreover, confidence has been found to change over time after the actual decision has been made (Yu et al., 2015;

Boldt & Yeung, 2015; Yeung & Summerfield, 2012). Therefore, in order to limit interference of postdecisional confidence processing, we required our participants to report their decision and the associated level of confidence simultaneously.

Our behavioral data show that performance did not depend on stimulus properties in our design and that, as we intended, (objective) stimulus difficulty was constant across trials. Still, participants reported higher confidence on some trials than on others and indeed, the accuracy across high-confidence trials was higher than across low-confidence trials. The latter suggests that they were able to evaluate their own performance and were aware of their own performance. If the external uncertainty in the stimuli is constant across trials but the subjects still experience a feeling of confidence which is predictive of their performance, then that suggests this confidence to be a readout of internal uncertainty.

4.2 Neural correlates of confidence

To investigate the neural systems involved in reading out confidence, we compared BOLD activation during high-confidence versus low-confidence responses. A number of brain areas exhibited larger BOLD responses with high-confidence than low-confidence decisions. First, the striatum was bilaterally activated, or more specifically, the head of the caudate nucleus and the adjacent nucleus accumbens, as well as a small portion of the left putamen. Second, some somatosensory and motor areas showed activation, namely the bilateral secondary somatosensory cortex (S2) and the left primary motor (M1), primary somatosensory (S1), and premotor (PMC) cortices. Below, I will discuss the possible roles of those areas in decision-making and perceptual confidence in more detail.

4.2.1 Striatum

We observed a robust bilateral activation around the head of the caudate nucleus and the nucleus accumbens, extending into the putamen in the left hemisphere only. The complex of caudate nucleus, nucleus accumbens, and putamen (together with the olfactory tubercle) is called the striatum. Functionally, the striatum as a whole is best known for its role in the planning of voluntary movements. The nucleus accumbens specifically, has often been referred to as the brain's reward center. More recently, it has become clear that the striatum plays a broader role in cognitive processes, specifically in goal-directed behavior and action selection (for reviews, see Grahn, Parkinson, and Owen, 2008 (caudate) and Floresco, 2015 (n. accumbens)).

The striatum is thus not typically thought to encode perceptual confidence and has not received much attention in the confidence literature so far. However, the nucleus accumbens and caudate nucleus have been associated with some other aspects of decision-making (Ding & Gold, 2010, 2013), and reward-related processes in particular. This area has been found to encode a prediction error on reward, i.e. the difference between expected and actual rewards (O'Doherty et al., 2004), as well as the amount of uncertainty associated with an upcoming reward (Preuschoff, Bossaerts, & Quartz, 2006). However, it appears that such responses are not driven by

reward per se or the anticipation thereof, but specifically occur when the (expected) reward is the outcome of an action performed by the subject (Tricomi, Delgado, & Fiez, 2004; Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004). Moreover, the reward does not need to be explicit; performance feedback on its own may be sufficient to elicit a reward-like response in the striatum, at least as long as participants are intrinsically motivated to improve their performance on the task (Tricomi, Delgado, McCandliss, McClelland, & Fiez, 2006). Thus, rather than responding to reward per se, the striatum seems to play a key role in learning action-outcome contingencies, where the outcome may be either an extrinsic or an intrinsic reward.

In our study, the same areas – nucleus accumbens, caudate – were found to respond to subjective confidence, even though participants did not receive any rewards, nor feedback on their performance. This finding is in agreement with two earlier studies in which stimulus properties were held constant (Hebart et al., 2016), or variability was controlled for in the analysis (Daniel & Pollmann, 2012). How should this similarity between the neural representations of confidence and reward be interpreted? One explanation could be that – provided that subjects are intrinsically motivated to perform well on the task – the feeling of being confident by itself is rewarding, especially in the absence of a stronger signal such as an external reward or feedback. The other way around, we can also argue that if a subject is required to make a decision which determines whether or not they will receive a reward, then their expectations of getting rewarded will depend on how confident they are that they have made the right choice. Thus, both subjective confidence and reward expectation are based on an evaluation of the quality of our own decisions and may, on a cognitive as well as a neural level, be more similar than they appear at first sight.

As for the role of the putamen in decision-making, it is more closely linked to the motor system than other areas within the striatum. In particular, the putamen is thought to be responsible for the execution of relatively automatic or habitual motor sequences and learning associations between stimuli and actions (Balleine, Delgado, & Hikosaka, 2007). Within the context of decision-making, it may thus be involved in the conversion of decisions into motor actions. This might also explain why we find activation of the left but not the right putamen here, as we found the proportion of right-hand responses to be significantly larger on high-confidence than on low-confidence trials. Whenever the subjects reported high confidence, the response was more frequently performed with the right hand than with the left, and therefore, averaged across all high-confidence trials, the left motor system was more active than the right.

We thus suggest that in the striatum – head of the caudate and nucleus accumbens in particular – the general quality of our perceptual decisions is evaluated, whether coupled to a reward or not. The putamen may be responsible for converting the confidence signal into an action. We hypothesize that the evaluation of the decision and computation of confidence is at least partially based on information about the amount of uncertainty in the internal representation of the stimulus at hand, although there may be other inputs, too.

4.2.2 Sensorimotor areas

Outside the striatum, we also observed some confidence-related activity in a number of somatosensory- and motor-related areas, namely bilateral S2 and left M1, S1, and PMC. The lateralization of sensorimotor activity to the left hemisphere might seem surprising at first sight, but can again be explained by the fact that we observed a correlation between the degree of confidence and the hand used to respond, with a bias towards right-hand responses for high-confidence trials and vice versa. Given the contralateral organization of both sensory and motor cortex (Fritsch & Hitzig, 1870; Penfield & Boldrey, 1937), it is then not surprising that confidence-related sensorimotor activity is more pronounced in the left hemisphere. Note that with a lower cluster-forming threshold ($p < 0.001$ instead of $p < 0.0001$) the same pattern of BOLD activity was found in the right hemisphere.

The question remains how confidence-related activity in sensorimotor areas should be interpreted. We believe this to be an artifact caused by differences in response finger and timing, which are – in the current design – correlated with the degree of the confidence. Given that in our task each confidence level (1–4) was paired with a different response button and therefore a different finger, and this mapping of confidence to fingers was the same across all sessions and all subjects, the average motor response and therefore also the corresponding somatosensory stimulation were slightly different for high- versus low-confidence judgments (Maldjian, Gottschalk, Patel, Detre, & Alsop, 1999; Olman, Pickett, Schallmo, & Kimberley, 2012). Besides the difference between high- and low-confidence trials with respect to the finger movement itself, the way in which the hand as a whole is stabilized during the button press also depends on which finger is used (Ivan Toni, personal communication). Moreover, we observed shorter response times for high confidence than for low confidence. Although response timing was corrected for in the GLM analysis, shorter response times suggest shorter motor preparation times, and thus potentially a more sudden increase of BOLD activity in motor areas. In a similar way, the somatosensory response could be affected, as faster motor responses might cause more sudden changes in somatosensory stimulation. Although the exact effects of variability in response and decision times on the motor and somatosensory signals have not yet been investigated properly, we interpret the observed activity in somatosensory and motor areas as a consequence of the variability in motor responses. First, response timing was correlated with the degree of confidence, and second, high- versus low-confidence judgments were coupled to different fingers, which required control over different sets of muscles for controlling the actual finger movement as well as stabilizing the hand as a whole.

4.2.3 Comparison to earlier confidence studies

In contrast to much of the earlier work on perceptual confidence (see section 4.1) we did not find confidence-related activity in parietal or (pre)frontal areas. What distinguishes these studies from ours, is that stimulus properties (e.g. contrast or motion coherence) were varied across trials. The problem with that, is that it enables subjects to use this variability as a predictor of their performance, with the risk that they might base their confidence

judgments on this external cue rather than on the amount of internal uncertainty. Therefore, we believe that these prefrontal areas may actually be involved in the monitoring of external stimulus properties rather than confidence assessment per se.

However, if indeed prefrontal areas are involved in monitoring the external noise (variability) in the stimulus, that does not necessarily mean that they do not contribute to some overall confidence variable computed in the striatum. Interestingly, the nucleus accumbens and (ventral) caudate nucleus appear to be functionally connected to prefrontal cortex and its orbitofrontal parts in particular (Di Martino et al., 2008). A possible explanation would thus be that the prefrontal cortex monitors properties of the external stimulus and sends output to the striatum which combines this with information from other sources to compute a joint confidence measure. In the absence of external noise or variability, the confidence measure would then be computed from other inputs to the striatum. Further research is required to investigate the functional interactions between these areas in the process of confidence estimation.

4.3 Future directions

This study provides new insights into the neural substrates of the computation of perceptual confidence in the absence of external noise, and suggests a key role for the striatum in reading out confidence from internal noise. However, further research is needed in order to confirm this hypothesis and to investigate whether this striatal representation of confidence is a universal one, i.e. whether activity in this area is predictive of confidence for other types of decisions or tasks as well. If so, then the question remains how exactly this confidence is computed. Given that most decisions are subject to multiple sources of noise, confidence is generally viewed as a summary variable. How the amount of uncertainty is estimated for each individual piece of information and how these uncertainties are integrated, both in terms of neural computations and areas involved, are topics for future research.

As for the specific paradigm used here, we suggest that confidence judgments were primarily based on a readout of the amount of noise in the cortical representation of the stimulus, as stimulus properties were kept constant across trials. It would thus be interesting to use the decoding methods recently developed in our lab to measure uncertainty in the stimulus representation in early visual cortex (van Bergen et al., 2015), and see how this uncertainty measure is propagated through the brain and how it relates to subjective confidence and to BOLD activity in the striatum.

Moreover, in this study, we found activity in several motor-related and somatosensory areas. We argue that this is mainly due to differences between high- and low-confidence responses with respect to the nature and timing of the motor response itself. However, we cannot entirely exclude the possibility that it is more than that, as it has previously been suggested that motor cortex may in fact be more than just an output stage and that (spontaneous) fluctuations in its activity may affect decision-making (Pape & Siegel, 2016). In our current design, we can hardly distinguish between

the sensory/cognitive and motor planning components of the decision process due to the correlations between motor response and confidence level. In follow-up studies we will thus aim to minimize such correlations by e.g. introducing a delay period before the response window in order to minimize reaction time differences between confidence levels, and counterbalancing finger-confidence level mapping across trials.

5 Conclusion

All sensory processing is subject to uncertainty: incoming sensory signals are inherently ambiguous due to noise in the external signal, and internal processing itself is affected by neuronal noise. This raises the question how we can infer from these noisy sensations the stimuli that caused them, and how we are able to make decisions based on those uncertain inputs. Current theories on human decision-making assume that our brains can estimate the amount of uncertainty associated with individual pieces of information and that, when making decisions, we weigh each piece of evidence according to its (estimated) degree of uncertainty. This idea is supported by experimental work.

If indeed the brain takes into account estimates of uncertainty while computing decision variables, then do we also have conscious access to this type of information? Humans are able to evaluate their own decisions in terms of confidence, but how and where in the brain subjective confidence measures are computed is not known. Here, we investigated correlations between BOLD activity and confidence reports in a visual orientation decision task. Note that stimulus properties were held constant across trials in this study. This was done to prevent that participants could generate expectations about their own performance based purely on features of the external stimulus. we show that humans experience different levels of decision confidence even in the absence of across-trial variability of stimulus properties. Moreover, confidence judgments were predictive of performance, suggesting that they were indeed based on some measure of internal uncertainty. As the fMRI data demonstrate, the primary locus of confidence-related BOLD activity in this task was the striatum, more specifically, the head of the caudate nucleus and the adjacent nucleus accumbens. Although previously, these areas have been associated with reward expectation more than confidence, we argue that these two concepts may be quite closely linked on a cognitive level since they both rely on an internal evaluation of the quality of our decisions. Thus, the striatum appears to play a key role in the computation or representation of perceptual confidence based on estimates of internal rather than external noise.

Bibliography

Andersson, J., Jenkinson, M., & Smith, S. (2007). *Non-linear regsitration, aka spatial normalisation. FMRIB technical report TR07JA2*. Retrieved from http://www.fmrib.ox.ac.uk/datasets/techrep/tr07ja2/tr07ja2.pdf

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychological Bulletin*, *78*(4), 266–278. doi:10.1037/h0033117

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27*(31), 8161–8165. doi:10.1523/jneurosci.1554-07.2007

Barthelmé, S. & Mamassian, P. (2010). Flexible mechanisms underlie the evaluation of visual confidence. *Proceedings of the National Academy of Sciences*, *107*(48), 20834–20839. doi:10.1073/pnas.1007704107

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., … Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152. doi:10.1016/j.neuron.2008.09.021

Boldt, A. & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478–3484. doi:10.1523/jneurosci.0797-14.2015

Brainard, D. H. (1997, January). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. doi:10.1163/156856897X00357

Daniel, R. & Pollmann, S. (2012, February). Striatal activations signal prediction errors on confidence in the absence of external feedback. *NeuroImage*, *59*(4), 3457–3467. doi:10.1016/j.neuroimage.2011.11.058

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. doi:10.1038/nn.3279

Di Martino, A., Scheres, A., Margulies, D. S., Kelly, A. M. C., Uddin, L. Q., Shehzad, Z., … Milham, M. P. (2008). Functional connectivity of human striatum: a resting state fMRI study. *Cerebral Cortex*, *18*(12), 2735–2747. doi:10.1093/cercor/bhn041

Ding, L. & Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *Journal of Neuroscience*, *30*(47), 15747–15759. doi:10.1523/jneurosci.2894-10.2010

Ding, L. & Gold, J. I. (2013). The basal ganglia's contributions to perceptual decision making. *Neuron*, *79*(4), 640–649. doi:10.1016/j.neuron.2013.07.042

Ernst, M. O. & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:10.1038/415429a

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–54. doi:10.1038/nn.2983

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*, 2811–2822. doi:10.1093/brain/awu221

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543. doi:10.1126/science.1191883

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annual Review of Psychology*, *66*(1), 25–52. doi:10.1146/annurev-psych-010213-115159

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, *7*(1), 30–40. doi:10.1006/nimg.1997.0306

Fritsch, G. & Hitzig, E. (1870). Ueber die elektrische Erregbarkeit des Grosshirns. *Archiv für Anatomie, Physiologie und Wissenschaftliche Medizin*, 300–332.

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141–155. doi:10.1016/j.pneurobio.2008.09.004

Greve, D. N. & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. doi:10.1016/j.neuroimage.2009.06.060

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: confidence in decision-making. *Neuroscience and Biobehavioral Reviews*, *55*, 88–97. doi:10.1016/j.neubiorev.2015.04.006

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639–1651. doi:10.1016/j.neuroimage.2003.11.029

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J. D. (2016). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex*, *26*(1), 118–130. doi:10.1093/cercor/bhu181

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002, October). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. doi:10.1006/nimg.2002.1132

Jenkinson, M. & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. doi:10.1016/S1361-8415(01)00036-6

Kepecs, A., Uchida, N., Zariwala, H. a., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. doi:10.1038/nature07200

Kiani, R. & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759–764. doi:10.1126/science.1169405

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Supplement), 14.

Knill, D. C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. doi:10.1016/j.tins.2004.10.007

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., & Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nature Neuroscience*, *16*(6), 749–755. doi:10.1038/nn.3393

Körding, K. P. & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. doi:10.1038/nature02169

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., & Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*, *84*(1), 190–201. doi:10.1016/j.neuron.2014.08.039

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. doi:10.1038/nn1790

Maldjian, J. A., Gottschalk, A., Patel, R. S., Detre, J. A., & Alsop, D. C. (1999). The sensory somatotopic map of the human hand demonstrated at 4 Tesla. *NeuroImage*, *10*(1), 55–62. doi:10.1006/nimg.1999.0448

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: from neural origins to behavior. *Neuron*, *88*(1), 78–92. doi:10.1016/j.neuron.2015.09.039

Middlebrooks, P. G. & Sommer, M. A. (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron*, *75*(3), 517–530. doi:10.1016/j.neuron.2012.05.028

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. J., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–454. doi:10.1126/science.1094285

Olman, C. A., Pickett, K. A., Schallmo, M.-P., & Kimberley, T. J. (2012). Selective BOLD responses to individual finger movement measured with fMRI at 3T. *Human Brain Mapping*, *33*(7), 1594–1606. doi:10.1002/hbm.21310

Orbán, G., Berkes, P., Fiser, J., & Lengyel, M. (2016, October). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, *92*(2), 530–543. doi:10.1016/j.neuron.2016.09.038

Pape, A.-A. & Siegel, M. (2016). Motor cortex activity predicts response alternation during sensorimotor decisions. *Nature Communications*, *7*, 1–13. doi:10.1038/ncomms13098

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi:10.1163/156856897X00366

Penfield, W. & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389–443. doi:10.1093/brain/60.4.389

Pleskac, T. J. & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. doi:10.1037/a0019737

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, *16*(9), 1170–1178. doi:10.1038/nn.3495

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. doi:10.1038/nn.4240

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, *51*(3), 381–390. doi:10.1016/j.neuron.2006.06.024

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neuroscience*, *3*(9), 940–945. doi:10.1038/78856

Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron*, *41*(2), 281–292. doi:10.1016/S0896-6273(03)00848-1

Tricomi, E. M., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, *18*(6), 1029–1043. doi:10.1162/jocn.2006.18.6.1029

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), 1728–1730. doi:10.1038/nn.4150

Vilares, I. & Körding, K. P. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. doi:10.1111/j.1749-6632.2011.05965.x

Volkmann, J. (1934). The relation of time of judgment to certainty of judgment. *Psychological Bulletin*, *31*(9), 672–673.

Yeung, N. & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B*, *367*(1594), 1310–1321. doi:10.1098/rstb.2011.0416

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489–510. doi:10.1037/xge0000062

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*, 509–517. doi:10.1016/S0896-6273(04)00183-7