

ACCEPTED MANUSCRIPT • OPEN ACCESS

Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision making

To cite this article before publication: Jacobo Fernandez-Vargas *et al* 2021 *J. Neural Eng.* in press <https://doi.org/10.1088/1741-2552/abf2e4>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2021 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence
<https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision making

Jacobo Fernandez-Vargas¹, Christoph Tremmel¹, Davide Valeriani^{2,3}, Saugat Bhattacharyya^{1,4}, Caterina Cinel¹, Luca Citi¹, Riccardo Poli^{1*}

¹Brain-Computer Interfaces and Neural Engineering laboratory, School of Computer Science and Electronic Engineering, University of Essex, Essex, UK

²Department of Otolaryngology | Head and Neck Surgery, Massachusetts Eye and Ear, Boston, MA, USA

³Department of Otolaryngology | Head and Neck Surgery, Harvard Medical School, Boston, MA, USA

⁴School of Computing, Engineering & Intelligent Systems, Ulster University, Londonderry, UK

*e-mail: rpoli@essex.ac.uk

Abstract. *Objective:* In many real-world decision tasks, the information available to the decision maker is incomplete. To account for this uncertainty, we associate a degree of confidence to every decision, representing the likelihood of that decision being correct. In this study, we analyse Electroencephalography (EEG) data from 68 participants undertaking eight different perceptual decision-making experiments. Our goals are to investigate (1) whether subject- and task-independent neural correlates of decision confidence exist, and (2) to what degree it is possible to build BCIs that can estimate confidence on a trial-by-trial basis. The experiments cover a wide range of perceptual tasks, which allowed to separate the task-related, decision-making features from the task-independent ones. *Approach:* Our systems train artificial neural networks to predict the confidence in each decision from EEG data and response times. We compare the decoding performance with three training approaches: (1) single subject, where both training and testing data were acquired from the same person; (2) multi-subject, where all the data pertained to the same task, but the training and testing data came from different users; and (3) multi-task, where the training and testing data came from different tasks and subjects. Finally, we validated our multi-task approach using data from two additional experiments, in which confidence was not reported.

Main results: We found significant differences in the EEG data for different confidence levels in both stimulus-locked and response-locked epochs. All our approaches were able to predict the confidence between 15% and 35% better than the corresponding reference baselines.

Significance: Our results suggest that confidence in perceptual decision making tasks could be reconstructed from neural signals even when using transfer learning approaches. These confidence estimates are based on the decision-making process rather than just the confidence-reporting process.

1 2 3 4 5 1. Introduction

6 7 *Decision-making*

8
9
10 A decision is the result of a process that integrates contextual cues and pre-existing
11 knowledge to commit to a categorical choice to achieve a particular goal. It has been
12 shown that during the decision-making process the human brain weighs and integrates
13 multiple noisy sources of information over time [1, 2, 3, 4, 5]. As a result, a meta-
14 cognitive evaluation of the decision is generated: the confidence [6, 7, 8], which reflects
15 the perceived probability of being correct and is generally correlated with the accuracy,
16 similarly to other behavioural and physiological measures, such as the response time
17 (RT) [9, 10, 11, 12, 13, 14, 15].
18
19

20 Different neural correlates of decision making have been identified using
21 neuroimaging techniques such as electroencephalography (EEG) [16, 17], including
22 neural correlates of confidence [18, 19, 8, 11, 20]. In particular, the activity in the
23 pre-frontal [21, 22] and parietal [23, 24] cortices correlates with the confidence reported
24 by human participants.
25
26

27 EEG has been used to characterise numerous brain states, such as mental workload,
28 valence, and arousal, which directly or indirectly affect decision making [25, 26, 27]. Also,
29 several studies have found differences in the event-related potentials (ERP) for different
30 confidence levels in decision making [28, 29, 5, 18, 20, 23]. These differences in the brain
31 activity make it possible to predict and classify confidence on a decision-by-decision
32 basis using machine learning algorithms [7, 30, 31, 20].
33
34

35 *Brain-computer interfaces*

36
37
38
39 Significant experience in the detection, prediction, and classification of trial-by-trial
40 brain responses has been acquired in the field of Brain Computer Interfaces (BCIs).
41 Normally, these devices are used to create a communication channel between a human
42 with significant motor disabilities and a machine [32]. The most common BCIs
43 record brain activity via EEG, thanks to its low cost, high temporal resolution, non-
44 invasiveness, and practicality. The downside of EEG is the low signal-to-noise ratio,
45 non-stationarity, and low spatial resolution. Due to these limitations, standard BCIs can
46 only issue a small set of commands and do so rather slowly and with occasional errors,
47 although the trade-off between speed and accuracy can often be adjusted [33, 34, 35].
48
49

50 BCIs can also be used for other forms of cognitive human augmentation [36]. For
51 instance, if EEG signals are complemented with behavioural and other physiological
52 recordings (Hybrid BCIs), one can obtain systems to support group decision-making
53 [37, 38, 39, 40, 41] that are capable of delivering significant practical benefits in real-
54 world situations.
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 *Transfer learning*

5
6 Due to high inter-subject variability of EEG signals, BCIs are usually required to be
7 trained with the brain signals of each user. The length of the training process limits BCI
8 applicability in many domains. Transfer learning could significantly reduce the training
9 needs of a BCI, by training the machine-learning model of the BCI with data from one
10 participant, and use the trained model with a different participant [42].

11
12 To date, transfer learning has only been used in established BCI applications, such
13 as ERP detection [43, 44], motor imagery classification [45, 46], and steady state visual
14 evoked potentials [47, 43]. This is, mainly, because of the non-stationarity of EEG, that
15 limits the efficiency of such approach. To the best of our knowledge, transfer learning
16 has not yet been applied to confidence decoding in decision making.

17
18 *Contributions*

19
20 This paper makes the following contributions.

21
22 *Neural correlates of confidence* . Most studies reporting analysis techniques for
23 confidence do it only for one task [48, 4, 6, 23, 7, 31, 12, 20, 21]. In this study, we instead
24 investigate the neural correlates of the confidence across *eight* different experiments
25 with 68 participants in total. This is a significative increase compared to classical
26 BCI experiments where 5 to 20 participants are used. In most of the cited papers, the
27 confidence levels are divided in two groups (Confident and non-Confident), either by
28 actively asking the participant only those two options, or calculating the median. We
29 grouped the confidence into four levels (Low, Mid, High, and Sure) to find whether
30 there was a gradient in the neural correlates associated to the confidence level. All
31 experiments included decision tasks where information was presented visually, but had
32 different stimuli and feedback. Analysing multiple experiments allowed to increase the
33 generality of results and interpretations. With this approach we expected to find both
34 tasks-related differences in the processes associated with confidence evaluation, as well
35 as a task-independent common biomarker of confidence. To the best of our knowledge
36 this is the first time an approach like this is been attempted.

37
38 *Confidence prediction* . The second goal of this study was to predict the reported
39 confidence. Of course, the best way to obtain the confidence is not to predict it but ask
40 directly the participant after each decision. However, being able to accurately predict
41 decision confidence is important in many time-critical tasks (e.g., in the military, in
42 trading, etc.) where waiting for people to evaluate and express their confidence is not
43 viable.

44
45 In previous studies, we focused on predicting the probability of each trial being
46 correct, as this is particularly useful for aiding group decisions [49, 39, 38, 40, 50, 51].
47 In this study however, we focus our attention on *predicting the confidence reported by*
48 *participants* after each decision. The rationale behind this choice is that, after task

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3 familiarisation, reported confidence tends to be a good estimator of performance [29].
4 Another advantage of estimating confidence rather than probability of correctness is
5 that the latter may not be readily available in every task to be able to train the BCI,
6 while the former can be asked at any time to the user during the training period.
7

8 Because confidence is essentially an analogue quantity, *we treat confidence*
9 *prediction as a regression problem with analogue outputs*, and do not require any prior
10 knowledge of the participant's confidence distribution compared to more traditional
11 approaches where the confidence prediction is treated as a binary problem [20, 7, 31]).
12 BCI systems are rarely used to solve regression problems. However, there do exist some
13 prior examples. For instance, the estimation of drowsiness, [52] reaction time, [53] or
14 hand position [54]. Nevertheless, to the best of our knowledge, a regression approach
15 has never been attempted for the prediction of the reported confidence.
16

17 Finally, for the first time, we investigated a zero-training approach to confidence
18 prediction. Zero-training is a form of transfer learning where the predictor is not
19 tailored for each participant. We used different approaches to predict the confidence
20 to investigate their impact on the quality of the prediction.
21

22 *Validation* . We validated our BCI confidence decoders using general zero training
23 in experiments where the participants did not report their confidence in the decisions.
24 This allowed us to investigate whether the neural correlates of confidence identified in
25 this study were related to the whole decision-making process, rather than being neural
26 correlates of confidence reporting.
27

36 2. Materials and methods

37 2.1. Experiments

38 For this study, eight different experiments conducted over the past 6 years were used. As
39 it can be seen in Figure 1, all experiments had similar trial structures: visual information
40 was provided, as either a static image or a video sequence. Following which, a decision
41 had to be made and reported by the participant together with their confidence (except
42 for the two validation experiments).
43

44 A brief description of each experiment is given below, while full details of each
45 experimental protocol are included in the corresponding publications.
46

47 *PATROL1* [50], *PATROL2*, and *PATROL3* In these three experiments, participants
48 were presented with a video of a corridor with doors at both sides. At a random
49 time, a soldier figure appeared for 250 ms, and participants had to decide (within a 2
50 s timeout) whether the figure was wearing a helmet or a cap, by pressing the left or
51 right mouse button, respectively. Participants were then asked to report the confidence
52 in their decision using the mouse wheel. A blue bar representing the confidence
53 varied accordingly as participants manipulated the wheel. If the participant did not
54

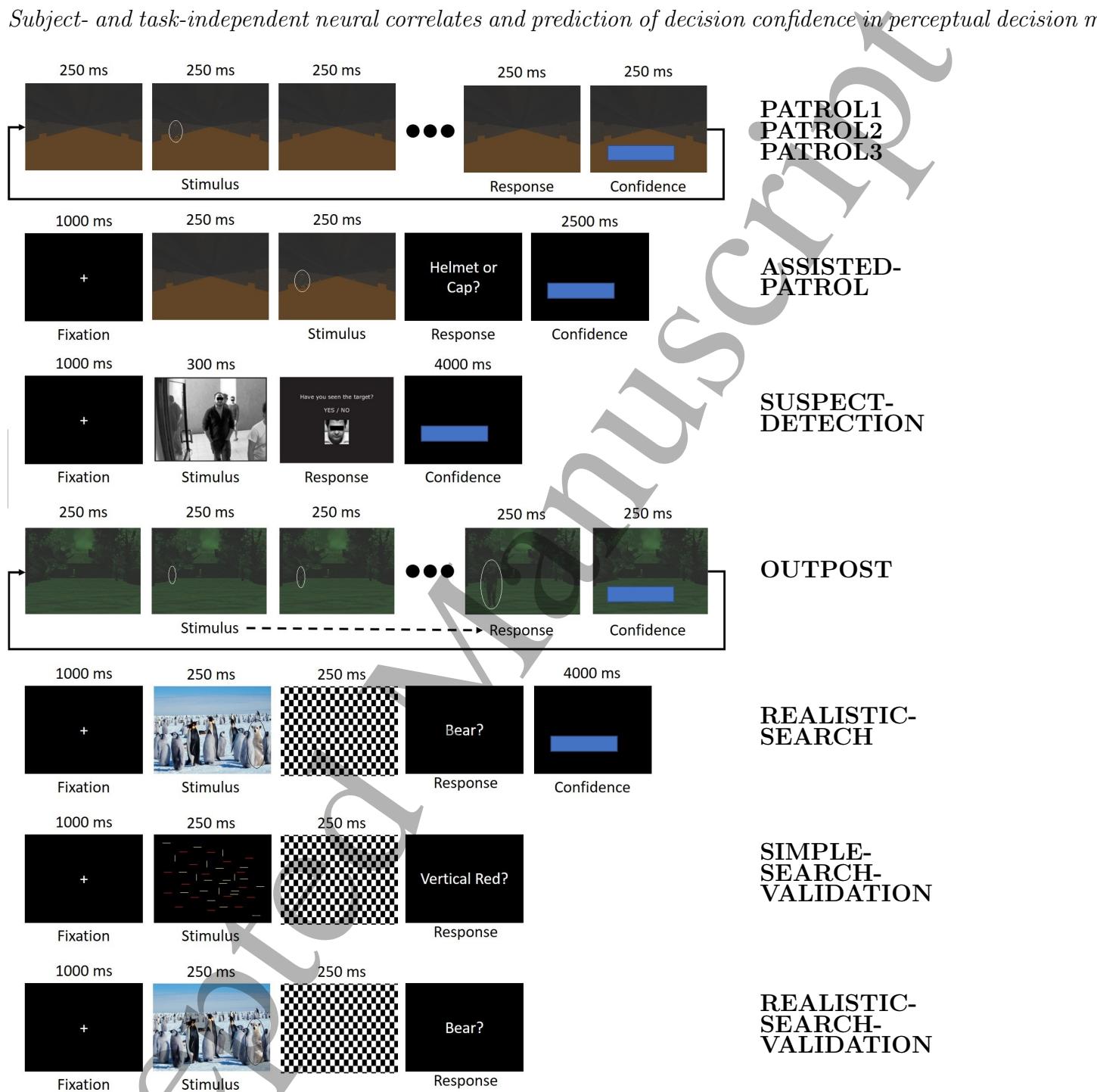


Figure 1. Trial structure for the experiments used in this article. In PATROL and OUTPOST, the arrow from the last to the first frame indicates that trials form a continuous sequence with no pauses, while the three black dots indicate that the system shows one frame every 250 ms (4 Hz). The dashed arrow in OUTPOST indicates that the stimulus changes over time. Finally, ellipses were superimposed to help the reader identify the character.

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 respond, the experiment would continue and their decision would be considered incorrect
5 (miss). In PATROL2 and PATROL3, after confidence was provided, feedback on the
6 correctness of the decision and confidence assessment was given to participants. In
7 PATROL2 feedback was in the form of a slider, which was most negative for incorrect
8 fully-confident decisions, weakly negative for incorrect low-confidence decisions, weakly
9 positive for correct low-confidence decisions, and most positive for correct fully-confident
10 decisions. In PATROL3, participants were shown their decision (represented by the
11 labels “Cap” and “Helmet”) and confidence (represented by a confidence slider) side
12 by side with the decision and confidence of an expert. Twelve participants took part
13 in PATROL1 experiment, while ten participants undertook PATROL2 and PATROL3.
14 Each participant performed 336 trials.

15
16 **ASSISTED-PATROL** This experiment, while being similar to the PATROL experiments,
17 presented some differences. Firstly, instead of video feeds, two static images
18 were presented (see Figure 1): (1) the empty corridor and (2) the corridor with a char-
19 acter, both displayed for 250 ms. Secondly, after the character image, participants were
20 forced to respond (i.e. there was no timeout for their responses). Thirdly, some of the
21 trials contained a cue before the stimulus indicating if the target would appear to the
22 right- or to the left-hand-side of the corridor. Finally, there was no trial-by-trial feed-
23 back. There were four types of trials: without any cue, with a static head that didn’t
24 provide any information, a voice saying “right” or “left”, or a talking head saying either
25 “right” or “left”. In this experiment, 12 participants were tested, each performing 640
26 trials, 160 per condition.*

27
28 **SUSPECT-DETECTION** [40] In each trial of this experiment, participants were
29 presented with a black-and-white photo of a crowded corridor (for 300 ms), and then
30 with a display asking if a specific person was present. They had to answer yes or no
31 by pressing the left or right mouse buttons, respectively, and then had to report their
32 confidence using the mouse wheel as in the PATROL experiments. In this experiment,
33 10 participants were tested, each performing 288 trials.

34
35 **OUTPOST** [51] In this experiment, participants were shown a video sequence
36 simulating the viewpoint of soldier at an outpost surveying a clearing. In each trial, a
37 character appeared in the distance and walked towards the outpost. Participants were
38 asked to decide as quickly as possible whether the character was wearing a helmet or

39
40 *The experiment was run (with the same protocol and amplifiers) jointly at the University of
41 Essex and at the University of Southern California. EEG was recorded from six participants (tested
42 at Essex) using a 64-electrode cap, while for the other six a 256-electrode EEG cap was used. The
43 256-electrode EEG dataset was down-sampled with bilinear interpolation to reconstruct EEG signals
44 from similar locations as for the 64-electrode cap. For four electrode locations (AF7, P9, AF8 and
45 P10) this was not possible, and the signals from these electrodes were discarded for all 12 participants.
46 Hence, preprocessing and data analysis for ASSISTED-PATROL were done using 60 electrodes.

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 a cap, reporting their decision using the mouse buttons. They then had to report the
5 confidence in this decision using the mouse wheel. The character stayed on screen until
6 the participant responded, so stimulus presentation time was not fixed. However, there
7 was a variable response time to make this decision. In this experiment 10 participants
8 were tested, each doing 360 trials.
9
10

11
12 *REALISTIC-SEARCH1 and REALISTIC-SEARCH2* [39] In these two experiments,
13 participants had to perform a visual search task. An image of an arctic environment
14 with a variable number of penguins and possibly a polar bear, photorealistically imposed
15 on the image, was presented for 250 ms. Participants had to decide whether there
16 was a polar bear or not in the image. Response and confidence were expressed
17 in the same way as in previous experiments. The difference between REALISTIC-
18 SEARCH1 and REALISTIC-SEARCH2 was that, in the latter, after reporting their
19 confidence, participants were informed about the confidence and decisions reported
20 by another participant, and had the possibility of changing their response. In the
21 REALISTIC-SEARCH1 experiment 10 participants were tested, while 16 took part in
22 the REALISTIC-SEARCH2 experiment. Each participant performed 320 trials.
23
24

25
26 *SIMPLE-SEARCH-VALIDATION* [38] This experiment was composed of 320 trials,
27 where participants were presented with a display composed of 40 green and red lines,
28 either vertical or horizontal, on a black background for 250 ms. Their task was to decide
29 whether or not there was a vertical red bar in the image by pressing the left or right
30 mouse buttons. Ten people took part in the experiment. This experiment was only used
31 to validate the models because the confidence was not reported.
32
33

34
35 *REALISTIC-SEARCH-VALIDATION* [49] The task was the same as for REALISTIC-
36 SEARCH1, except that participants were not asked to report their confidence. Ten
37 subjects took part in the experiment, each performing 320 trials. This experiment was
38 not used for training, but only to validate the models in situations where confidence was
39 not reported.
40
41

42 In all experiments, confidence was reported using a scale from 0 to 1 in steps of 0.1,
43 creating 11 possible reported confidence values.
44
45

46 2.2. Exclusion criteria 47

48
49 Both the raw signals and the descriptive statistics of the EEG recordings were inspected
50 to determine the quality of the recordings. This resulted in the removal of three
51 participants from PATROL1 and one from PATROL2. Furthermore, we excluded the
52 participants with an accuracy more than two standard deviations lower than the average
53 accuracy across all participants. These participants were likely not paying attention to
54 the experiment or did not understand the task. As a result, six participants with
55 accuracy lower than 56% were excluded: two from SUSPECT-DETECTION, one from
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3

4 REALISTIC-SEARCH1, and three from REALISTIC-SEARCH2. Moreover, because
5 we were focusing our analysis on the confidence, we removed those participants for which
6 the distributions of confidence in the correct trials and in the incorrect trials were not
7 significantly different, as assessed by the Wilcoxon rank-sum test resulting in $p > 0.05$.
8 After this process another 12 participants were excluded: one from PATROL2, one from
9 SUSPECT-DETECTION, three from OUTPOST, one from REALISTIC-SEARCH1,
10 and six from REALISTIC-SEARCH2. Therefore, data from a total of 68 participants
11 were included in the analysis.
12
13

14 We also removed the trials in which the participants reported a 0 confidence, as
15 this measure was used by participants in different ways. Some would use it to indicate
16 that they had responded randomly, while others used it to indicate a wrong decision. In
17 total, these trials represented only 1.7% of the total. Full details about the distribution
18 of the accuracy, RT, and proportion of each confidence level can be found on 1 , 2, and
19 3 respectively in the additional materials.
20
21

22 *2.3. Setup and preprocessing*
23
24

25 In all experiments, participants sat comfortably at about 80 cm from an LCD screen
26 while wearing an EEG cap connected to a Biosemi ActiveTwo system. All the
27 experiments were performed using wet electrodes and, except for six participants in
28 ASSISTED-PATROL, all the recordings were performed with 64 electrodes in the
29 standard 10-20 system.
30
31

32 The EEG data were preprocessed as described in [39]. In brief, the original data
33 was sampled at 2048 Hz, then band-pass filtered between 0.15 and 40 Hz using an FIR
34 filter. Then, the signal was down sampled by a factor of 16, resulting in a 128 Hz signal.
35 In addition to this, a correction for eye-blink and other ocular movements was performed
36 using a subtraction algorithm based on correlations to the average differences between
37 FP1 and F1 and between Fp2 and F2 [55].
38
39

40 After the preprocessing, two types of epochs were extracted from EEG in each
41 trial: stimulus-locked and response-locked. The former started at stimulus onset and
42 lasted for 2.5 s, to ensure inclusion of the response and its neural correlates in every
43 experiment. Response-locked epochs started 1.25 s before the response and lasted for
44 1.5 s, as we were mostly interested into the neural processes leading to a decision. A
45 baseline was calculated and removed for each epoch and channel. In each epoch, the
46 mean of the signal from 25 ms before and 25 ms after the stimulus or response onset was
47 used as baseline correction. Finally, when performing ERP analyses, an epoch rejection
48 process was applied. For each participant, the difference between the maximum and
49 minimum voltage of each epoch was computed. Then, a threshold was set as the third
50 quartile plus 1.5 times the difference between the first and the third quartiles. The
51 epochs that had a difference between the maximum and the minimum voltage higher
52 than the threshold were removed from the analysis, on a channel-by-channel basis.
53
54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4

5 2.4. *Neural correlates of decision confidence*

6
7 We grouped trials into four categories: *Low confidence*, for trials with a reported
8 confidence between 0.1 and 0.3; *Mid confidence*, for trials with a reported confidence
9 between 0.4 and 0.6; *High confidence*, for trials with a reported confidence between 0.7
10 and 0.9; and *Sure*, for trials with a reported confidence of 1. The reason to have an
11 independent category for *Sure* is that participants report full confidence much more
12 frequently than any other value of confidence (32% of the trials overall). We chose fixed
13 boundaries for the confidence compared to other studies because this allowed use of
14 the same methodology across participants and tasks. Having boundaries based on the
15 percentiles of the reported confidence allows to have a somewhat calibrated confidence,
16 as well to have a more balanced set of labels for the classification. However, to do this
17 its necessary to have the distribution of the confidence of the user a-priori.
18
19

20 Two methods were used to analyse the EEG activity for different confidence levels.
21 First, we performed a *single-experiment analysis*, where we calculated the average epoch
22 voltage across trials for each confidence level (without grouping by subject), as well as
23 the 95% confidence interval.
24
25

26 Next we performed a *multi-task analysis*, where we first calculated the epoch average
27 of each participant and confidence level. Then the grand average from those averages
28 was derived. Participants that did not have at least 10% of trials at each confidence level
29 were removed from this analysis. This resulted in three participants being removed: one
30 from SUSPECT-DETECTION and two from REALISTIC-SEARCH2.
31
32

33 Statistical analyses were conducted using the non-parametric Kruskal-Wallis test
34 comparing the average voltages between confidence levels for each time point and
35 channel.
36
37

38 2.5. *Confidence regression and transfer learning*
39

40 The system was designed with the following structure:
41
42

- 43 (i) Feature extraction: Uses the raw EEG data and RT as input for each trial (32769
44 values) and returns the calculated features (129).
- 45 (ii) Feature selection: Using a greedy algorithm based on linear regression, the features
46 that do not contribute positively to the prediction are removed.
- 47 (iii) Predictor: Using the selected features, an ANN with the same number of inputs as
48 the number of features selected, and 10 neurons in the output layer is used. The
49 activations of the output layer are weighted to obtain the final prediction.
50
51

52 In this section we will describe each of these items one by one. The three approaches
53 used for different levels of transfer learning were implemented following the upcoming
54 structure.
55
56

57 2.5.1. *Features* Two features for each EEG channel were selected from each trial, one
58 from the stimulus-locked epoch and another from the response-locked epoch. For the
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 stimulus-locked epochs we used the ERP amplitude, calculated as the mean voltage
5 between 500 and 750 ms after stimulus onset minus the mean voltage in the preceding
6 500 ms. For the response-locked epochs, we used the mean amplitude of the EEG
7 between 1250 and 500 ms before the response. Combining these features for all channels
8 we had 128 neural features (120 for the ASSISTED-PATROL experiment) for each trial.
9

10
11 Additionally, we used the response time as a feature, since it has been demonstrated
12 that both confidence and correctness are correlated with RT in a variety of situations [56,
13 14, 57, 24, 37].
14

15 To investigate the contribution that the RT may have over the prediction, we
16 performed the training and testing using only the RT as an input for the classifier.
17

18
19 *2.5.2. Feature selection* The method used to select the features was an iterative greedy
20 algorithm that first measured the prediction accuracy with a set of features. Then it
21 removed one feature at a time, re-calculating the prediction accuracy to determine how
22 much this changed when the feature was removed. Then the algorithm permanently
23 discarded the two features that, when removed, changed prediction accuracy the least.
24 At that point the algorithm started again, repeating the process with the remaining
25 features until the desired number of features remained. This algorithm is an adaptation
26 of an algorithm described in previous work [58]. Here, when we calculated prediction
27 accuracy, we trained and tested (with five-fold cross-validation) a linear regression model
28 using the available features. We used linear regression instead of an ANN for reducing
29 the computational burden of this optimisation procedure.
30

31
32 *2.5.3. Predictor* The first step of the prediction was an ANN that followed a shallow
33 network approach with 10 neurons in the output layer. The ANN was designed to predict
34 the confidence from the selected features with the aforementioned algorithm. This meant
35 that different models had different input sizes. Each feature was z-scored corrected, as
36 this is standard for most machine learning systems. This method prevents features with
37 a higher mean and variation from becoming more relevant to the classification than
38 they should be. The network had 10 output neurons, one for each of the ten possible
39 confidence levels (from 0.1 to 1 in steps of 0.1). A dropout layer (with a probability of
40 0.5) was used to reduce overfitting [59]. The network had one hidden layer, the neurons
41 of which used a hyperbolic-tangent activation function. Finally, a softmax layer was
42 used to sharpen the network outputs.
43

44 Cross validation was used to decide the number of neurons in the hidden layer.
45 The values tested were 5, 10 and 20. The maximum number of training epochs was set
46 to 500, with a mini batch size of 10% of the data. In addition to this, an early-stop
47 criterion based on the error of a validation set was used. A fraction (20%) of the trials
48 were extracted, without replacement, from the training set to form the validation set.
49 Training was stopped after six epochs in which the error on the validation set did not
50 decrease.
51

52 The i -th output of the softmax layer for a particular input pattern was taken to
53

54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 represent the probability, w_i , of confidence level $cl_i \in \{0.1, 0.2, \dots, 1\}$ being reported in
5 a trial, for all i 's. However, instead of using the most probable confidence level as the
6 predicted confidence, we used a weighted sum of the most probable confidence levels.
7 The goal of this was to obtain an analogue output, not limited to just 10 different
8 confidence values. To predict the confidence in a trial, the N most probable confidence
9 levels cl_i 's were selected. Assuming the reorder the softmax outputs by w_i , w_1 being
10 the largest, w_2 the second largest and so on, N was set trial by trial to be the smallest
11 number for which $\sum_{i=1}^N w_i \geq 0.5$.

12 * Then, the predicted confidence for that trial was calculated using the following
13 formula:

14
$$\frac{\sum_{i=1}^N w_i * cl_i}{\sum_{i=1}^N w_i}. \quad (1)$$

15
16 2.5.4. *Prediction approaches* To assess the efficacy of a zero-training method, we tested
17 three different approaches. These approaches were validated using cross validation, while
18 the cross validation used for the feature selection was nested inside:

19
20 *Single Subject (SS):* A model was trained and tested for each participant individually,
21 using standard cross validation. This non-zero-training approach provided a baseline
22 reference level of performance to compare the other two approaches.

23
24 *Multi Subject (MS):* A transfer-learning model across subjects was trained and tested
25 separately for each experiment, using a *leave-one-subject-out* cross-validation approach.
26 The results from this model represented the performance of a system that does not know
27 anything about the future user, but is likely specialised to the task performed in each
28 experiment. This approach was a form of transfer learning across participants.

29
30 *Multi Task (MT):* In this generalised model with transfer learning across tasks, the
31 training and test sets followed a *leave-one-experiment-out* cross-validation approach.
32 This model represented a fully-generalised approach, where the system is independent
33 from both the subject and the perceptual decision-making task.

34
35 2.5.5. *Evaluation methods* For each of these approaches, we used an Artificial Neural
36 Network (ANN) as predictor (details are in a later section). As a baseline for the ANN
37 we selected the prediction error made by a classifier that always predicted the mean of
38 the reported confidence (this baseline is more conservative than using just random). To
39 evaluate the different approaches we used the following metrics: (1) the median absolute
40 error between the predicted confidence and the reported confidence (MAE), (2) the

41
42 *Setting this threshold to 0.5 was a compromise between the need to provide a more graded
43 confidence prediction than using cl_1 (i.e., the most probable output of the ANN) and the need to
44 be able to predict values at the extreme of the range of the confidence scale (e.g., 1, which is the most
45 common response).

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 prediction's median meta-cognitive accuracy, (3) the prediction's mean confidence delta
5 ($c\Delta$), and (4) the prediction's mean confidence calibration. We defined (2)–(4) below.

6
7 The *meta-cognitive accuracy* (MCA) is a quantity that indicates how good a person
8 is at evaluating their own decisions. [60] There are different methods of measuring it
9 [61, 29] with their own advantages and disadvantages. Here, we computed the MCA as
10 follows:

$$11 \quad MCA = 1 - |confidence - correctness| = \begin{cases} confidence & \text{if } correctness = 1, \\ 1 - confidence & \text{otherwise.} \end{cases} \quad (2)$$

12
13 where the correctness is 0 (incorrect) or 1 (correct), and the confidence is a value
14 between 0 and 1. Therefore, the MCA takes values between 0 and 1, with 1 representing
15 a correct decision made with the highest confidence or an incorrect decision made with
16 0 confidence, and 0 representing an incorrect decision made with the highest confidence
17 or an correct decision made with 0 confidence.

18
19 The formula above applies to each individual decision. We were however, interested
20 in aggregate statistics, such as $E[MCA]$ for a participant across all trials in an
21 experiment.

22
23 The limitations with this measure are that the MCA is heavily influenced by the
24 difficulty of the task and may be influenced in a counter intuitive manner by confidence
25 biases (such as those seen in over confident or under confident individuals). For instance,
26 in a relatively easy task (where participants make infrequent mistakes) a participant
27 could obtain a high average MCA by just responding always with the highest value of
28 confidence, irrespective of whether the decision was correct or incorrect.

29
30 To complement the meta-cognitive accuracy, we also defined a new measure, the
31 *confidence delta* or $c\Delta$, which represents to what extent the confidence recorded in
32 correct decisions is higher than that recorded in incorrect decisions:

$$33 \quad c\Delta = E[confidence|correctness = 1] - E[confidence|correctness = 0]. \quad (3)$$

34
35 A value of $c\Delta = 1$ would indicate that the participant reported full confidence to all
36 the correct trials and a 0 confidence to all the incorrect ones. A $c\Delta = 0$ would indicate
37 that reported confidence is random or that the same confidence is reported for every
38 trial. We calculated this value on a subject-by-subject basis.

39
40 Finally, another desirable property of the confidence (either reported or predicted)
41 is *calibration*. Confidence is calibrated if, on average, it matches the probability of
42 decisions being correct. To quantify the degree of calibration we defined the following
43 *calibration offset*:

$$44 \quad \text{calibration offset} = |E[confidence] - P(correctness = 1)|. \quad (4)$$

45
46 This value reflects how close the predicted confidence is to the accuracy of the
47 participant, a positive *calibration offset* indicating for instance that a participant is
48 either overconfident or under-confident.

49
50 When taken together, these metrics provide a comprehensive evaluation on the
51 quality of the our confidence predictors.

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4
5
6
7
8
9
10
11
12
13

14
15 *2.5.6. Statistical Analysis* For each one of the four evaluation methods, first we performed a Wilcoxon test to compare the ANN and the Baseline methods. Then, we performed a Kruskal-Wallis analysis using only the ANN methods. If this resulted in a p-value < 0.05, we then performed three Wilcoxon pair test. In those cases where a difference was found, the effect size was calculated using Cohen's δ . Assuming a large effect size (.5), with an α 0.05 and a Power (1 - β) 0.8, the sample size for a two tail t-test is 26.

16
17 *2.6. Model validation*

18 As mentioned before, four measures were used to assess the quality of the prediction:
19 MAE, MCA, $c\Delta$ and calibration offset.

20 To further validate the Multi Task models, we used two additional experiments
21 (SIMPLE-SEARCH-VALIDATION and REALISTIC-SEARCH-VALIDATION; see
22 “Experiments” section and Figure 1) as validation data. The confidence for those
23 experiments was predicted using the trained models. The difference with the training
24 experiments was that participants were not asked to report their confidence in the
25 validation experiments.

26 To predict the confidence in these two experiments, a new ANN was trained with the
27 Multi Task approach, but cross validation was not used. Instead all eight experiments
28 were used for training, using the number of neurons in the hidden layer that resulted in
29 the lowest error during the training of the MT models.

30 To evaluate performance in the validation experiments, the MCA, $c\Delta$, and
31 calibration offset were calculated using the outputs of the ANN on the two validation
32 experiments. It was not possible to calculate the MAE given that there was no ground
33 truth to compare the predictions with. This validation method allowed us to test
34 whether the neural correlates observed and, thus the prediction made, arose from the
35 decision-making process or were linked to the confidence-reporting process.

36
37 **3. Results**

38 In this section, the results of the study are reported. These are divided into three areas:

39 Neural correlates of confidence decision, in which we analysed the neural correlates
40 of confidence in a large dataset of eight different experiments and 68 participants.
41 Furthermore, we divided the confidence level into four groups to get a finer grain analysis
42 compared to the typical confident vs non-confident categorisation.

43 Confidence prediction and transfer learning, where we used a BCI system to predict
44 the confidence in single trials. Using BCIs to predict continuous values is not common;
45 in particular, we could not find any study predicting confidence in decision-making. In
46 addition, we investigated the feasibility of using transfer learning to build a system that
47 can be used in a plug-and-play manner.

48 Model validation, where, exploiting the transfer leaning capabilities of our system,

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 we went a step further and predicted the confidence of decision-making tasks where
5 participants were not asked to report their confidence. This validation provides further
6 support for the results and shows how generalizable the proposed method is.
7
8

9
10 *3.1. Neural correlates of decision confidence*

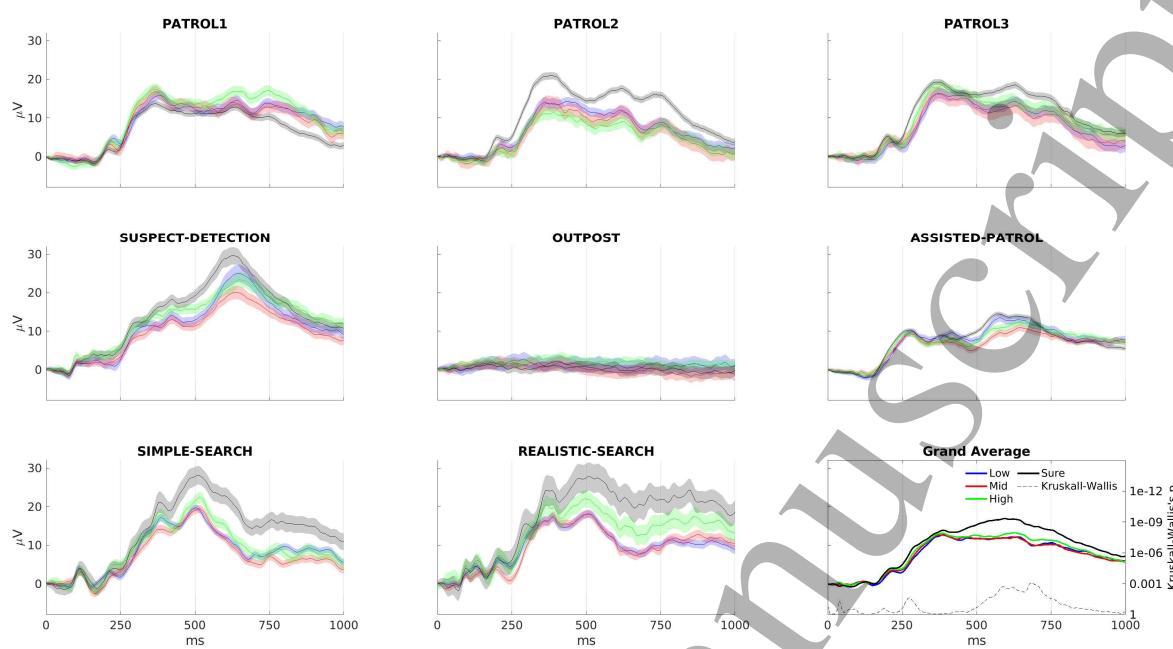
11
12 *3.1.1. ERPs* Figure 2 shows the results for stimulus-locked epochs and channel Pz as
13 the parietal area has previously shown to correlate with the confidence during perceptual
14 decision making [23]. In most experiments we see differences in the averages recorded for
15 different confidence levels at 350–600 ms after the stimulus onset. Also, ERPs recorded
16 in the 'Sure' condition were, in most tasks, higher than in the other levels of confidence.
17 The PATROL1 and OUTPOST experiments were exceptions to this behaviour. In the
18 case of the OUTPOST experiment, we could not see any real ERPs. This is because,
19 due to the nature of the experiment design, the appearance of the character on the
20 display can be quite difficult to detect at first, and participants might take a relatively
21 long time from the appearance of the stimulus before they are ready to respond. Hence,
22 there is a large variability in ERPs latencies across trials and participants, resulting
23 in flat averages. In the case of the PATROL1 experiment, ERPs were present but the
24 stimulus-locked epochs have no difference between the confidence bins.
25
26

27 Figure 3 shows the results for response-locked epochs. Here, the differences between
28 the four confidence levels are much more visible, and the ERPs' morphology seems
29 to vary proportionally with the confidence level. Using response-locked epochs, it is
30 possible to observe differences between classes in PATROL1, which were not visible in
31 the stimulus-locked epochs. However, for the OUTPOST experiment we still did not
32 measure any difference between the confidence classes.
33
34

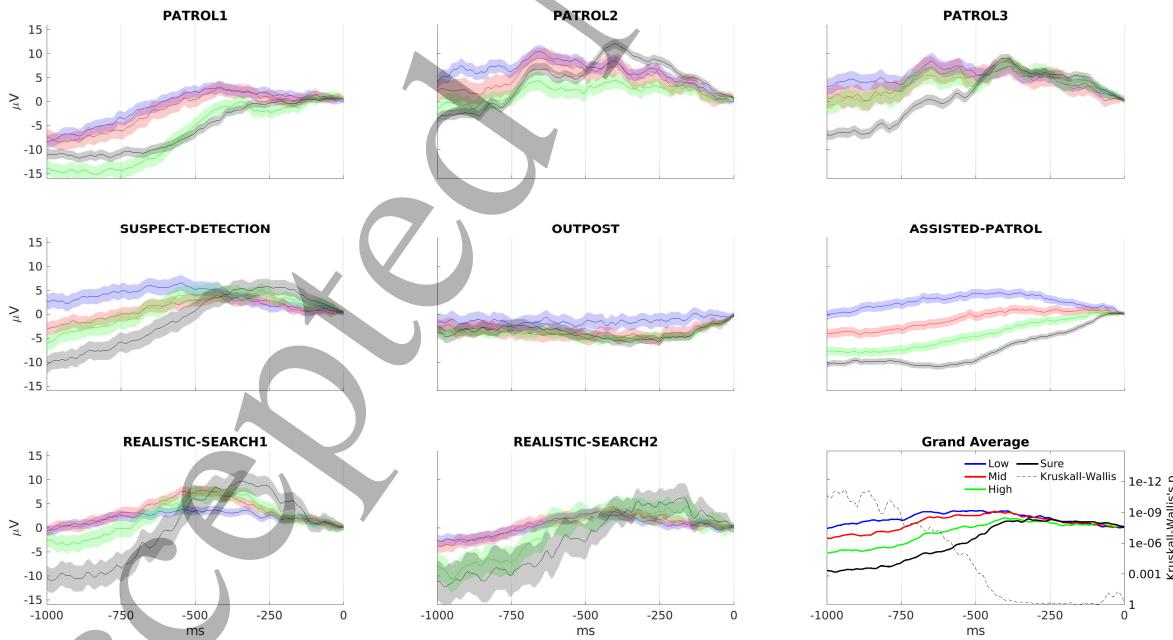
35 *3.1.2. Grand averages across experiments* The bottom right plots in Figures 2 and 3
36 show the grand average across all experiments for channel Pz, and the *p*-value of the
37 Friedman test (logarithmic ordinate scale on the right-hand side of the plot) for stimulus-
38 and response-locked epochs, respectively. The 'Sure' class is significantly different from
39 the grand-averages of the other classes for stimulus-locked epochs only, between 600 and
40 750 ms after the stimulus onset. On the contrary, the response-locked grand-averages
41 show differences between all four confidence levels, particularly up until around 500 ms
42 before the response.
43
44

45 Figure 4 shows the Kruskal-Wallis *p*-values across scalp locations and time for
46 stimulus- and response-locked epochs. The test verifies the hypothesis that ERPs
47 recorded in the four confidence classes are not drawn from the same distribution.
48 Statistical differences among the four classes are lateralised in the left hemisphere, and
49 are present in many channels for both stimulus- and response-locked representations.
50 Differences are stronger between 550 and 750 ms after stimulus presentation in the
51 stimulus-locked epochs. Differences are even stronger in response-locked epochs, as we
52 expected from our earlier observations on the bottom right plots in Figures 2 and 3.
53
54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3



26 **Figure 2.** Mean and 95% confidence interval of EEG activity at electrode Pz in
27 *stimulus-locked* epochs for the four confidence levels grouped by experiment. The
28 grand average plot (bottom right) shows the average ERPs across experiments and the
29 p -value of the Friedman test (logarithmic ordinate scale on the right-hand side of the
30 plot).



53 **Figure 3.** Mean and 95% confidence interval of EEG activity corresponds to the
54 *response-locked* epochs for the four confidence levels grouped by
55 experiment. The grand average plot (bottom right) shows the average ERPs across
56 experiments and the p -value of the Friedman test (logarithmic ordinate scale on the
57 right-hand side of the plot).

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 To confirm that the differences observed in the grand averages were not a result of
5 differences in correctness rather than confidence levels, we split the data between correct
6 and incorrect trials and performed the same analysis. The result of such analysis can
7 be seen in Figure 7 in the additional materials. For correct trials (80% of the total) we
8 can observe very similar results to the obtained in the original analysis. This by itself
9 confirms that the differences observed for confidence levels are not due to the correctness
10 of the response. In the case of incorrect trials (20% of the total), the differences are
11 only notable for response locked trials, where the effects are magnified by the difference
12 RT between confidence levels. It is important to note that, due to the exclusion criteria,
13 only 32 participants were valid for this analysis.

14 Finally, to control the effect of the RT, we performed an analysis on the response
15 locked epochs grouping by RT. We created three intervals based on the RT in seconds:
16 [0.5, 0.9), [0.9, 1.2), [1.2, 2). These thresholds were chosen so that, overall, there were
17 the same number of trials for each condition. However, due to the exclusion criteria, this
18 resulted in only 47, 53, and 28 valid subjects being included for each condition, down
19 from the 65 available in the original analysis. The goal was to remove the variance
20 of the RT from the analysis and confirm that the differences in the confidence were
21 not just due to the relative temporal shifts in ERPs induced by response locking. The
22 plots can be seen in Figure 8 in the Additional Materials. The results show that the
23 differences in the response locked epochs are noticeable, in particular for trials falling in
24 the RT intervals [0.5,0.9) and [0.9,1.2) which are reasonably narrow. However, we found
25 no differences for trials in the interval [1.2,2). This confirms that, even if there is a
26 correlation between RT and confidence, not all the confidence variance can be explained
27 by differences in RT.

28
29 *3.2. Confidence regression and transfer learning*

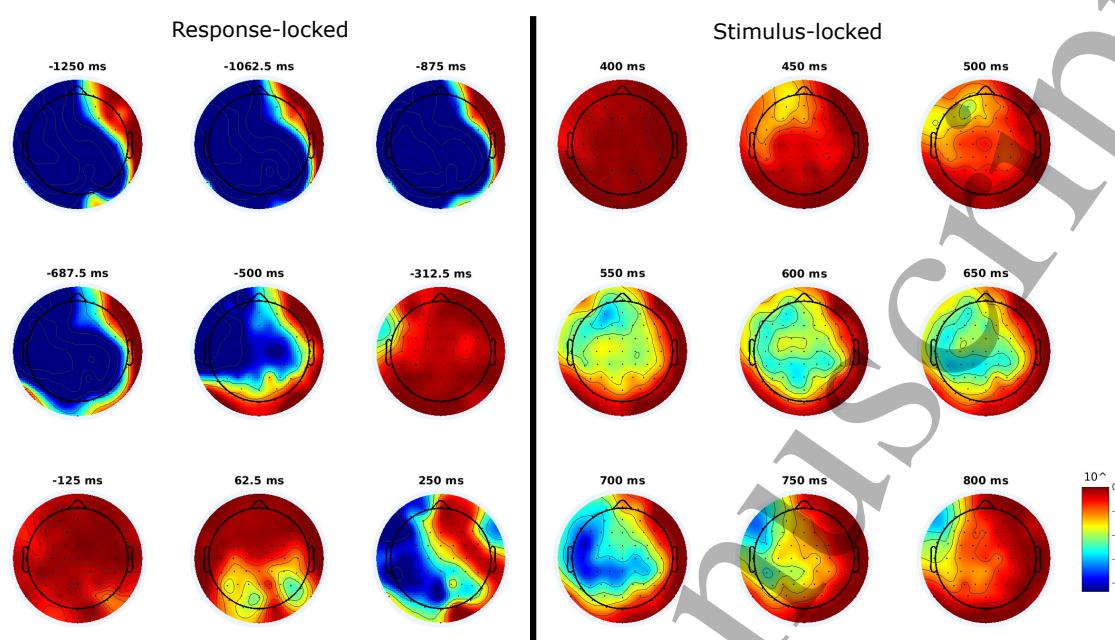
30 Figure 5 shows the average of the four variables across subjects used to evaluate the
31 quality of the prediction. Furthermore, in Figures 9, 10, 11, and 12 the values for
32 each subject, and task are presented for the MAE, meta cognitive accuracy, $c\Delta$, and
33 calibration offset respectively. The Figure 5.A shows the median absolute error (MAE)
34 between the predicted confidence and the reported confidence across all subjects for the
35 three approaches (SS, MS, and MT) and methods (Baseline and ANN) being compared.
36 As expected, the MAE is minimal for the SS approach, slightly worse for MS and even
37 worse for MT. It also appears that Baseline is worse than NN.

38 The MAE of ANN methods was significantly lower than the MAE of the
39 corresponding baseline method (Wilcoxon $p < 0.001$). Moreover, the MAE of the
40 three ANN-based approaches were significantly different (Kruskal-Wallis $p < 0.001$).
41 Finally, pairwise Wilcoxon rank-sum tests on the ANN-based approaches validated that
42 SS ANN had lower MAE than MS approach, which in turn had lower MAE than the
43 MT approach (Bonferroni-corrected $p < 0.002$).

44 Additionally, we compared the three approaches when only the RT was used

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3



25 **Figure 4.** Scalp maps of Kruskal-Wallis's p-values at different times for both stimulus-
26 locked and response-locked epochs. Note that the colour scale is logarithmic. The
27 p-values in the response-locked epochs have been clipped to be comparable to the ones
28 in the stimulus-locked epochs.
29

30
31 compared to when both the EEG and RT were used. The results showed that using
32 only the RT leads to better results (MAE) for the SS approach (0.109 vs 0.139), but
33 worse for MS (0.213 vs 0.209) and MT (0.282 vs 0.272) approaches. In all cases, the
34 pairwise Wilcoxon rank-sum tests resulted in a Bonferroni-corrected $p < 0.001$. These
35 differences are even more clear if we look at accuracy (weight of the diagonal in the
36 confusion matrix). This value was better for the only RT method in the SS approach
37 (32% vs 36%) but worse for the MS (24% vs 16%) and MT (13% vs 9%) approaches.
38
39

40 Figure 5.B shows the mean MCA for SS, MS and MT with the Baseline and ANN
41 predictors. The chart also reports the ground truth value ('Real') obtained when using
42 the confidence reported by the participant and the one obtained using the validation
43 data set ('Val.'). We performed the same statistical tests for the MCA as we did for
44 the MAE. All three Wilcoxon rank-sum tests comparing ANN and Baseline for different
45 prediction strategies (SS, MS and MT) resulted in p -values < 0.001 , indicating that the
46 ANN achieved a better MCA than the Baseline. The Kruskal-Wallis test comparing the
47 three ANN approaches returned a p -value of 0.057, indicating that the three approaches
48 were not statistically different. Because the p -value is (marginally) non significant, we
49 did not perform the corresponding pairwise comparisons. Finally, we further performed
50 three Wilcoxon rank tests to compare the MCA from the ANN (with the SS, MS and
51 MT approaches) to the ground truth value. The Bonferroni-corrected p -values for the
52 three comparisons were 0.095, 1, and 0.029 for SS, MS, and MT, respectively. This
53 indicates that the MCA was significantly worse than the real value only in the MT
54 approach.
55
56

57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4 approach.
5

6 Figure 5.C shows the results for $c\Delta$ for ANN values (as $c\Delta = 0$ for the baseline) as
7 well as ground-truth ('Real') and the validation data set ('Val.'). As it can be seen from
8 the figure, $c\Delta$ of the ANN is much smaller (about 1/4 for MT, and 1/3 for SS and MS)
9 than for the ground-truth (the actual reported confidence). Similarly to the previous
10 performance metrics, we first performed a Kruskal-Wallis test comparing the three ANN
11 approaches. The Kruskal-Wallis test indicated a significant difference between the three
12 ANN approaches ($p = 0.005$). Then, we performed three paired Wilcoxon tests to
13 compare the different approaches. This resulted in the Bonferroni-corrected p -values 1,
14 0.018, and 0.012 for SS vs MS, SS vs MT, and MS vs MT, respectively, indicating that
15 the MT approach had a significantly lower $c\Delta$ than the other two methods. Additionally,
16 $c\Delta$ was significantly lower for the reconstructed confidence compared to the ground truth
17 (Wilcoxon $p < 0.001$). Finally, the $c\Delta$ distribution was significantly different from zero
18 ($p < 0.001$ for the three approaches).
19
20

21 Figure 5.E shows the results of the $c\Delta$ value for individual experiments for the
22 MT approach. This helps us to visualise that, even if the standard error of the $c\Delta$ for
23 the ground-truth is more or less similar across various experiments, the mean varies
24 significantly across them. However, the mean $c\Delta$ of the reconstructed confidence is
25 more stable. It is also interesting to note that there seems to be no correlation between
26 $c\Delta$ values calculated from the ground-truth and the ones calculated from the predicted
27 confidence ($\rho^2 = 0.005$, $p = 0.867$)
28
29

30 Finally, the mean calibration offset is shown in Figure 5.D. The confidence
31 reported by participants ('Real') seems less calibrated than the predictions of the ANN,
32 particularly for the MT ANN. As before, the first analysis that we performed was a
33 Kruskal-Wallis analysis to see if there was any difference between the three approaches.
34 This resulted in a p -value of 0.145, suggesting that there was no difference between
35 them. We then compared each of the approaches with the ground truth, which resulted
36 in Bonferroni-corrected p -values of 1, 0.218, and 0.106 for SS, MS, and MT respectively.
37 This indicates that, in terms of calibration offset, the predicted confidence was similar
38 to the real one.
39
40

41 3.3. Model validation 42

43 Validation results are represented by the green bars in Figure 5. For validation
44 purposes, we reconstructed the (missing) confidence values using an MT ANN approach.
45 As discussed previously, it was not possible to compute MAE between the reported
46 confidence and the reconstructed one because the confidence was not reported in the
47 validation experiments. For each of the three remaining performance metrics, we
48 performed a Wilcoxon rank test to see whether the performance measures obtained
49 with the reconstructed confidence values were statistically different from those obtained
50 with MT ANN approach on the eight training experiments. Both the comparisons of
51 the MCA and $c\Delta$ resulted in $p < 0.001$, while the comparison of the calibration offsets
52 with the MCA resulted in $p = 0.005$.
53
54

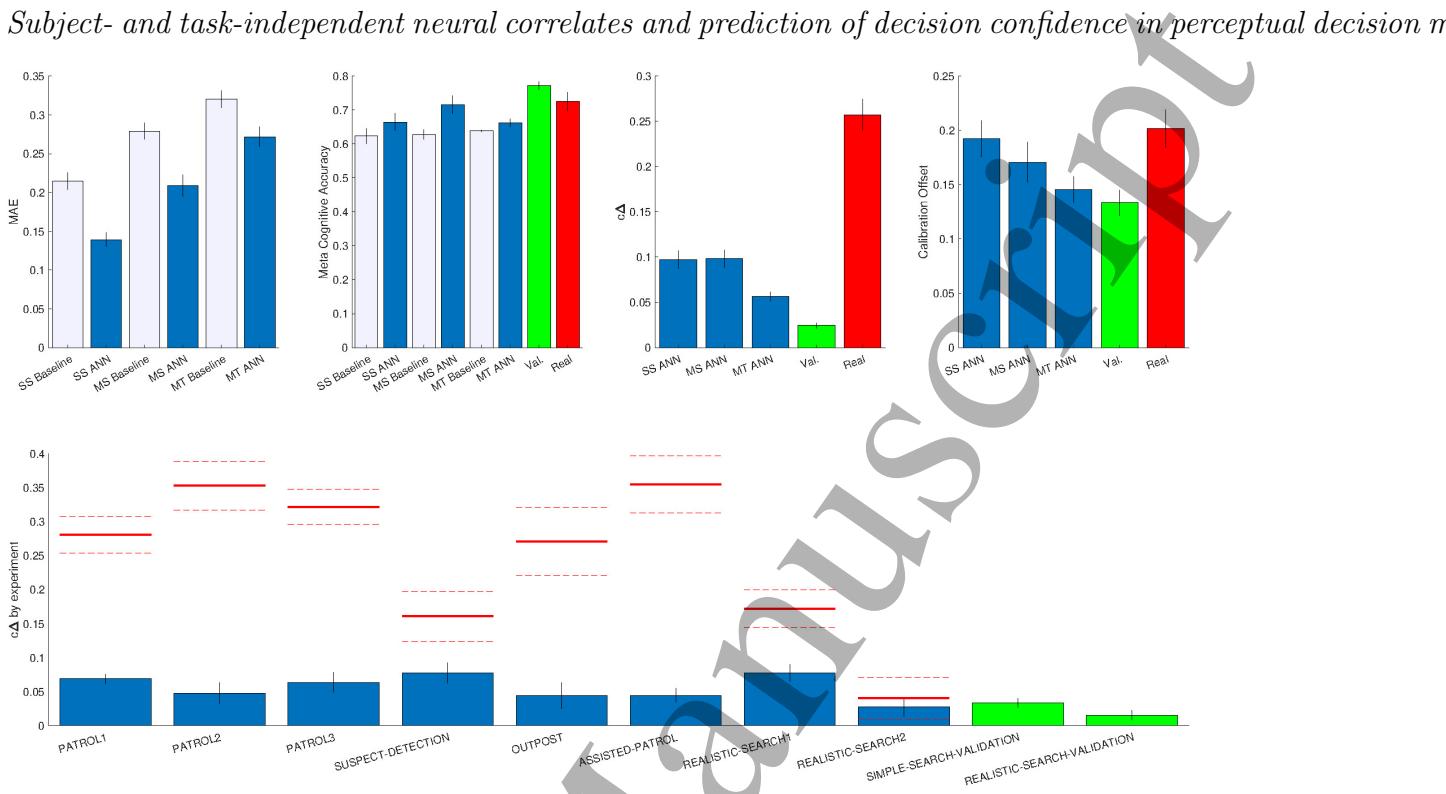


Figure 5. Evaluating the quality of confidence predictions. The four bar charts at the top of the figure show the mean of each measure evaluated across all subjects, while the bar chart at the bottom shows the $c\Delta$ (how different the confidence is between correct and incorrect trials) results for each experiment separately (also averaged across subjects). In the top bar charts, dark blue bars represent the results for the ANN method, light blue bars represent the baseline results, green bars correspond to the validation sets and red bars are the results obtained when the real reported confidence was used. In the bottom plot, red continuous lines indicate the mean of the reported confidence, dotted lines indicate the standard error, and the blue bars show $c\Delta$ for MT predictions. In all bar charts, the vertical black lines indicate standard error of the corresponding data.

resulted in $p = 0.687$. These results indicate that the validation data performed similarly to the MT ANN approach in terms of calibration offset, better in terms of MCA and worse in terms of $c\Delta$. As with the training data, we performed a Wilcoxon rank-sum test on the $c\Delta$ values to test whether the median was equal to zero. This resulted in a p-value < 0.001 . Finally, we tested whether the difference in confidence medians between correct and incorrect responses were statistically significant. This resulted in 48 (70.6%), 48 (70.6%), and 43 (63.2%) participants with significantly different confidence means between correct and incorrect trials for SS, MS, and MT approaches, respectively. Performing the same analysis over the 20 participants of the validation data set, showed that 12 (60%) had a significantly different means.

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 **4. Discussion**

5
6 In this study, we instead investigate the neural correlates of the confidence across *eight*
7 different experiments with 68 participants in total. Additionally, instead of dividing
8 the confidence into only two groups (Confident and non-Confident), we grouped the
9 confidence into four levels (Low, Mid, High, and Sure) to find whether there was a
10 gradient in the neural correlates associated to the confidence level.
11
12

13
14 *4.1. Neural correlates of decision confidence*

15 As we have seen in Figures 2 and 3 different ERPs are associated with different levels of
16 confidence. This is clearer in response-locked epochs than in stimulus-locked ones. In
17 many practical applications the event that caused a response is not known *a priori*, but
18 the response is, of course, always known. Our results suggest that even if the information
19 about the stimulus onset is not available, it may still possible to obtain strong neural
20 correlates of confidence using response-locked epochs.

21 Overall, the results indicate that there is a significant degree of similarity in the
22 ERPs associated with different levels of confidence in many experiments. For instance,
23 the higher the confidence the higher the ERP amplitude in several experiments with
24 stimulus-locked epochs and in most experiments with response-locked epochs.* In
25 the case of response-locked epochs, this can also be seen in the grand average across
26 experiments on the bottom right of Figure 3.

27 In particular, in several experiments, the condition 'Sure' was associated with
28 significantly bigger ERPs than the other confidence levels in both stimulus- and
29 response-locked epochs. This is visible in both grand averages (in Figures 2 and 3).
30 Reporting the confidence requires: deciding the precise confidence to report. This can
31 be easy and quick when the target is comfortably recognised, however, for less confident
32 decisions, deciding the precise confidence to report may involve lengthier and more
33 complex processes.

34 The time at which the differences between ERPs peak in stimulus-locked epochs,
35 varies slightly from experiment to experiment, but tends to be between 500 and 750 ms
36 after stimulus presentation. This is slightly later than what was found in [6] but it is
37 similar to the interval reported in [20]. The differences with the former study could be
38 attributed to differences in tasks and stimuli.

39 The low *p*-values in the statistical test observed in the *Pz* grand averages and scalp
40 maps, show that there are significant differences in the ERPs associated to different
41 confidence levels that are common across experiments. For this reason such ERPs should
42 represent the mental processes associated with any (perceptual) decision making. The
43 *p*-value scalp-maps in Figure 4 show that ERPs associated with different confidence
44 values start being statistically different at 550 ms after stimulus onset in the frontal
45

46 *The response-locked epochs are base-line corrected at the response. So, here ERP amplitude is
47 judged by comparing the minimum and the maximum voltage recorded.
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 area as found in [21]. Then, differences remain lateralised but spread to the parietal
5 area, similar to what was reported in [23]. The lateralisation may be due to the fact
6 that most participants in our experiments were right handed.
7
8

9 The *p*-value scalp-maps in Figure 4 show that also the corresponding ERPs for
10 response-locked epochs are lateralised, but also spread to central areas. These differences
11 are most likely due to the high correlation between response time and confidence.
12 For this reason, trials where decisions were made with higher confidence, have shorter
13 response times than trials where participants were less confident.
14
15

16
17 4.2. *Confidence regression and transfer learning*
18

19 In this study we were able to predict the confidence in a continuous way with better
20 accuracy than the baseline (average reported confidence). This was true even for the
21 approaches where the method was built with data from completely different subjects
22 (MS), or different subjects and experiments (MT). We can see from Figure 5 and from
23 the statistical analyses performed, that, from the point of view of the MAE, the best
24 approach to predict the confidence is, unsurprisingly, SS followed by MS and MT.
25 However, from the point of view of the meta-cognitive accuracy and the calibration
26 offset, the three approaches were not statistically different. For the $c\Delta$ measurement, the
27 MT approach was inferior to the SS and MS approaches, but these were not statistically
28 different from each other. This means that even if the accuracy of the prediction is
29 better for the SS approach than the MS, their predictions have the same separability
30 between correct and incorrect trials in terms of predicted confidence.
31
32

33 Furthermore, in Figure 6 we can see that for both the SS and MS approaches, the
34 ANN method predicts values across all the range of possibilities, with a distribution
35 similar to the real one, a skewed distribution with many 'Sure' cases and an uniform-
36 like distribution for the rest of the confidence levels. On the other hand, given that
37 the baseline only predicts one value (the mean confidence) for each fold, the confusion
38 matrices for the baseline show white spaces. For the MT approach, most of the ANN
39 predictions are around the mean reported confidence, like the baseline, with a wider
40 distribution similar to a Gaussian. Considering all these results, it appears that the
41 MS approach is the most promising of the three. Even if it lacks the generalisation of
42 MT, it is still a zero-training approach. This provides significant advantages towards
43 creating "plug-and-play" decision support systems, where the confidence of a participant
44 could be predicted straight away. This could be used, for instance, to evaluate when
45 the decision maker may need a break. Additionally, the confidence prediction from the
46 SS and MS approach are equally good as surrogates of the accuracy.
47
48

49 Regarding the $c\Delta$ value obtained in different experiments (see bottom of Figure 5),
50 we observed a large standard deviation in every experiment, not only for the MT
51 approach, but also for the ground truth. Interestingly, the predicted $c\Delta$ was not
52 correlated with the real $c\Delta$ (p -value=0.255), nor with the accuracy (p -value=0.051).
53 However, there was a correlation between the real $c\Delta$ value and the accuracy (p -
54
55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

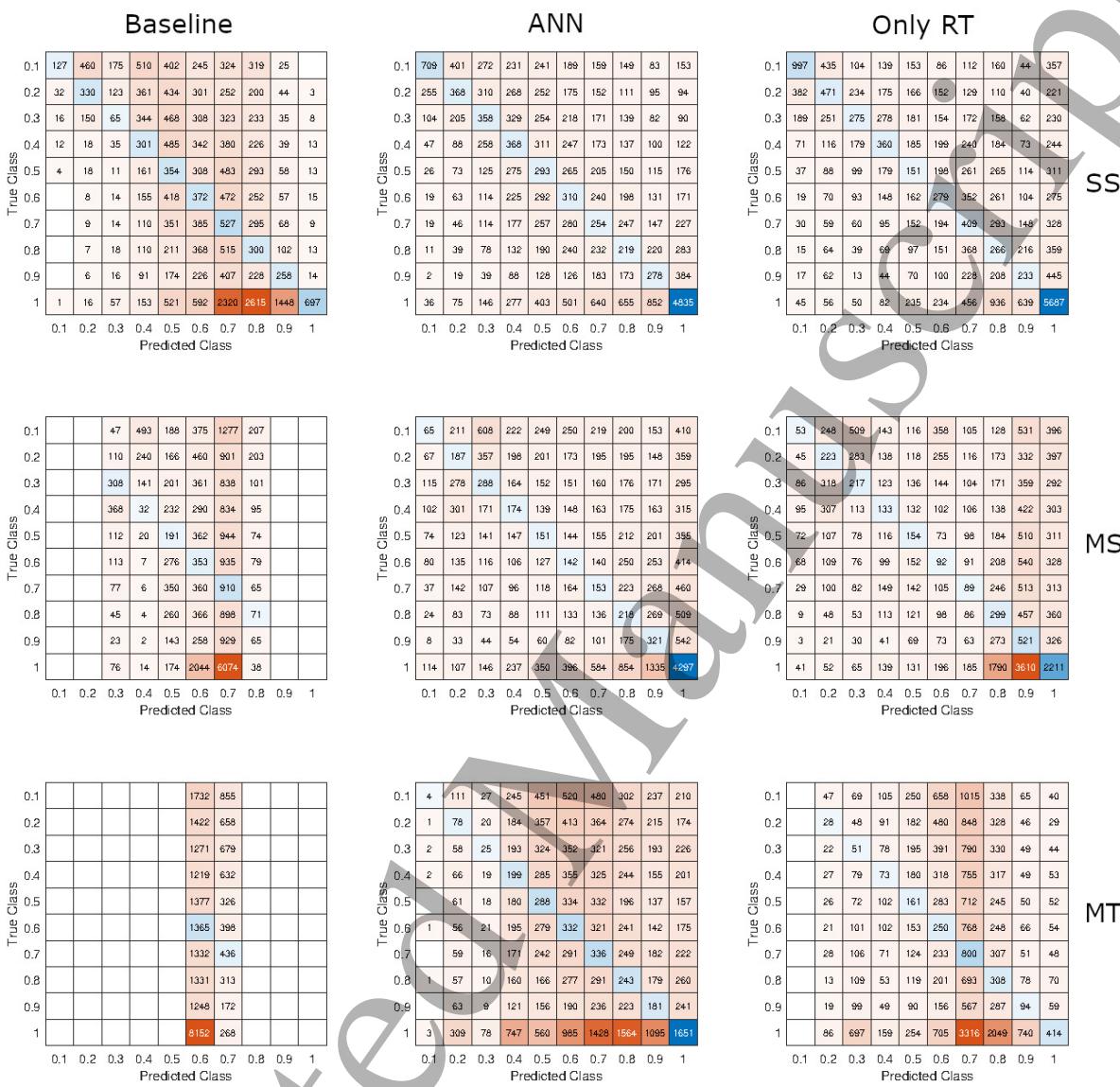


Figure 6. Confusion matrices for the three methods, Baseline, ANN, Only RT and three approaches, SS, MS and MT. The diagonal of each matrix correspond to the correctly predicted confidence values, while the other cells are incorrectly predicted confidences. The number in each cell indicates the number of trials falling in the corresponding class. The stronger the colour (red for incorrect and blue for correct) the more trials in a cell.

value<0.001). This suggest that the ability of the subject and, thus, the system to properly separate the correct and incorrect trials is directly impacted by the participant's accuracy.

Finally, considering the confusion matrices in Figure 6, and the results when comparing the ANN method with the RT-Only, it appears that using only RT is sufficient for the SS approach, as it shows little variation within task and subject. However, it seems that the EEG helps to generalise the prediction which is needed for the MS and MT approaches. In particular, it can be seen in the confusion matrix for MT, that

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 the prediction of the RT-Only method shows smaller variability on the prediction. For
5 instance, the latter never predicts the 0.1 confidence level.
6
7
8

9 4.3. Regression models validation
10

11 Given that in both experiment SIMPLE-SEARCH-VALIDATION and REALISTIC-
12 SEARCH-VALIDATION, the confidence was not reported, it is not possible to be certain
13 of the quality of the prediction. Nevertheless, the meta-cognitive accuracy of MT shows
14 levels comparable to the ones obtained in the SS and MS approach, suggesting the
15 accuracy might be similar to the one obtained with the 8 training problems for which
16 confidence was reported. This result supports the hypothesis that the confidence that
17 can be detected through ERPs is not linked to the confidence-reporting process, but
18 instead, to the actual decision-making process. This opens up the possibility of obtaining
19 the users' confidence without having to ask for it directly after each decision. In addition,
20 these results reinforce all the previous studies about neural correlates of confidence,
21 suggesting that the features discovered in those studies were actually related to the
22 decision-making process too. However, we cannot be certain whether the predicted
23 confidence in the validation problems is equally useful as surrogate of the confidence
24 like in the others experiments. Even if the statistical analysis indicated that the $c\Delta$ was
25 significantly larger than 0, if we look at the results of individual experiments (bottom of
26 Figure 5), we can see that these results vary significantly in different experiments (not
27 only for the validation data, but for the training data as well). For example, SIMPLE-
28 SEARCH-VALIDATION shows a $c\Delta$ similar to REALISTIC-SEARCH, OUTPOST, or
29 ASSISTED-PATROL. However, REALISTIC-SEARCH-VALIDATION shows $c\Delta$ values
30 that are not statistically different from 0. More experiments will be needed to find
31 conclusive answers.
32
33
34
35
36
37
38

39
40 5. Conclusions
41
42

43 In this study we have shown that there is a difference in the ERPs elicited during the
44 decision-making process for different confidence values. In particular, those trials where
45 the participants answered that they were sure of their decision, showed the biggest
46 differences. This points to the idea that there is more difference between being certain
47 of a decision and having any degree of uncertainty, than the difference between different
48 levels of uncertainty. This should be taken into consideration for the design of future
49 experiments. Instead of allowing subjects a fine grain range of confidence levels, broader
50 categories (one of them being 'Sure') show more separability in the EEG signals.
51
52
53

54 We have shown that some of the differences across confidence levels are present
55 across experiments. Additionally, we have shown that it is possible to predict the
56 confidence using only EEG signals and the response time, with better accuracy than
57 the baseline. This is not only true for single subject approaches, but also for zero
58 training approaches such as predicting one participant's confidence using the rest of the
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3
4 participants as training data, or even more, predicting the confidence of every participant
5 of a specific experiment trained with other experiments. It is important to note that
6 some of the experiments tested had notable differences in the stimulus presentation
7 and one even had a different number of electrodes. Even if the MAE is higher for
8 the zero-training approaches compare to the single subject, the MS approach was not
9 significantly different from the SS approach in terms of: MCA, $c\Delta$, and calibration
10 offset. In the future, further investigation should be done on the limits of the transfer
11 learning capabilities by, for example, testing more and more different experiments, or
12 by removing part of the data. Also, the model should be improved in the future. In this
13 study a shallow ANN was used. However deep networks have shown their potential on
14 the field [62, 63, 64] making them a good candidate for future research on confidence
15 prediction.

16 Finally, we further validated the model built by predicting the confidence in two
17 experiments where the confidence was not reported. The predicted confidence had a
18 $c\Delta$ significantly different from zero. This demonstrated that the predicted confidence
19 was linked to the decision-making process rather than the confidence-reporting process.
20 We realise that making a validation using a data from different experiments may not be
21 always possible. However, we consider that this kind of validation is a step forward to
22 have BCI systems used in less controlled environments and in real-world applications.

23
24
25
26
27
28
29
30
31
32 **6. Acknowledgments**

33
34 **Disclaimer:** This article is an overview of UK MOD sponsored research and is released
35 for informational purposes only. The contents of this article should not be interpreted
36 as representing the views of the UK MOD, nor should it be assumed that they reflect
37 any current or future UK MOD policy. The information contained in this article cannot
38 supersede any statutory or contractual requirements or liabilities and is offered without
39 prejudice or commitment. The Authors acknowledge support of the UK Defence Science
40 and Technology Laboratory (Dstl) and Engineering and Physical Research Council
41 (EPSRC) under grant EP/P009204/1. This is part of the collaboration between US
42 DoD, UK MOD and UK EPSRC under the Multidisciplinary University Research
43 Initiative. DV acknowledges support from the US DoD Bilateral Academic Research
44 Initiative (BARI) under grant W911NF1810434. We would also like to acknowledge
45 and thank the contribution of Morgan Mason for the revision of this text.

- 46
47
48
49
50
51 [1] Gold J I and Shadlen M N 2007 *Annual Review of Neuroscience* **30** 535–574 ISSN 0147-006X (Preprint NIHMS150003) URL
52 <http://www.annualreviews.org/doi/10.1146/annurev.neuro.29.051605.113038>
53
54 [2] de Lange F P, Jensen O and Dehaene S 2010 *Journal of Neuroscience* **30** 731–738
55
56 [3] Cheadle S, Wyart V, Tsetsos K, Myers N, De Gardelle V, Castañón S H and Summerfield C 2014
57 *Neuron* **81** 1429–1441
58
59 [4] Pisauro M A, Fouragnan E, Retzler C and Philiastides M G 2017 *Nature Communications* **8** 15808
60 ISSN 20411723
61 [5] Polánía R, Krajbich I, Grueschow M and Ruff C C 2014 *Neuron* **82** 709–720 ISSN 10974199

- 1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
- 3
4 [6] Graziano M, Parra L C and Sigman M 2015 *Journal of Cognitive Neuroscience* **27** 1090–1103 ISSN
5 15308898
- 6 [7] Krumpe T, Gerjets P, Rosenstiel W and Spüler M 2020 *Brain-Computer Interfaces* ISSN 23262621
- 7 [8] Kepecs A, Uchida N, Zariwala H A and Mainen Z F 2008 *Nature* **455** 227–231
- 8 [9] Adler W T and Ma W J 2018 *PLoS Computational Biology* **14** ISSN 15537358
- 9 [10] Sanders J I, Hangya B and Kepecs A 2016 *Neuron* **90** 499–506 ISSN 10974199
- 10 [11] Faivre N, Filevich E, Solovey G, Kühn S and Blanke O 2018 *Journal of Neuroscience* **38** 263–277
11 ISSN 15292401
- 12 [12] Grimaldi P, Lau H and Basso M A 2015 There are things that we know that we know, and there
13 are things that we do not know we do not know: Confidence in decision-making
- 14 [13] Ratcliff R and Sterns J J 2013 *Psychological Review* **120** 697–719 ISSN 1939-1471
- 15 [14] Baranski J V and Petrusic W M 1994 *Perception & Psychophysics* **55** 412–428 ISSN 0031-5117
- 16 [15] Aitchison L, Bang D, Bahrami B and Latham P E 2015 *PLoS Computational Biology* **11** ISSN
17 15537358
- 18 [16] Christopoulos G I, Tobler P N, Bossaerts P, Dolan R J and Schultz W 2009 *Journal of Neuroscience*
19 **29** 12574–12583
- 20 [17] Philiastides M G and Sajda P 2006 *Cerebral cortex* **16** 509–518
- 21 [18] Boldt A and Yeung N 2015 *Journal of Neuroscience* **35** 3478–3484
- 22 [19] Molenberghs P, Trautwein F M, Böckler A, Singer T and Kanske P 2016 *Social cognitive and*
23 *affective neuroscience* **11** 1942–1951
- 24 [20] Gherman S and Philiastides M G 2015 *NeuroImage* **106** 134–143 ISSN 10959572
- 25 [21] Gherman S and Philiastides M G 2018 *eLife* **7** ISSN 2050084X
- 26 [22] Basten U, Biele G, Heekeren H R and Fiebach C J 2010 *PNAS* **107** 21767–21772
- 27 [23] Herding J, Ludwig S, von Lautz A, Spitzer B and Blankenburg F 2019 *NeuroImage* **201** 116011
28 ISSN 10959572
- 29 [24] Kiani R and Shadlen M N 2009 *Science (New York, N.Y.)* **324** 759–64 ISSN 1095-9203
- 30 [25] Wang X W, Nie D and Lu B L 2014 *Neurocomputing* **129** 94–106
- 31 [26] Berka C, Levendowski D J, Lumicao M N, Yau A, Davis G, Zivkovic V T, Olmstead R E, Tremoulet
32 P D and Craven P L 2007 *Aviation, space, and environmental medicine* **78** B231–B244
- 33 [27] Tremmel C, Herff C, Sato T, Rechowicz K, Yamani Y and Krusinski D J 2019 *Frontiers in Human*
34 *Neuroscience* **13**
- 35 [28] Ojanen V, Revonsuo A and Sams M 2003 *Psychophysiology* **40** 192–197 ISSN 0048-5772 URL
36 <http://doi.wiley.com/10.1111/1469-8986.00021>
- 37 [29] Yeung N and Summerfield C 2012 *Philosophical Transactions of the Royal Society B: Biological*
38 *Sciences* **367** 1310–1321
- 39 [30] Vi C and Subramanian S 2012 Detecting error-related negativity for interaction design *Proceedings*
40 *of the SIGCHI Conference on Human Factors in Computing Systems* pp 493–502
- 41 [31] Kubanek J, Hill J, Snyder L H and Schalk G 2015 *Frontiers in neuroscience* **9** 243
- 42 [32] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 *Clinical*
43 *neurophysiology* **113** 767–791
- 44 [33] Müller-Putz G, Schwarz A, Pereira J and Ofner P 2016 From classic motor imagery to complex
45 movement intention decoding: The noninvasive Graz-BCI approach *Progress in Brain Research*
46 (Elsevier) pp 1–32
- 47 [34] Birbaumer N 2006 *Psychophysiology* **43** 517–532 ISSN 00485772
- 48 [35] Vaid S, Singh P and Kaur C 2015 EEG signal analysis for BCI interface: A review *International*
49 *Conference on Advanced Computing and Communication Technologies, ACCT* vol 2015-April
50 (Institute of Electrical and Electronics Engineers Inc.) pp 143–147 ISBN 9781479984886 ISSN
51 23270659
- 52 [36] Cinel C, Valeriani D and Poli R 2019 *Frontiers in human neuroscience* **13** 13
- 53 [37] Poli R, Valeriani D and Cinel C 2014 *PloS one* **9**
- 54 [38] Valeriani D, Poli R and Cinel C 2017 *IEEE Transactions on Biomedical Engineering* **64** 1238–1248
- 55
56
57
58
59
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*

3 ISSN 1558-2531

- 4
- 5
- 6 [39] Valeriani D, Cinel C and Poli R 2017 *Scientific Reports* **7** 7772 ISSN 2045-2322 URL
7 <https://doi.org/10.1038/s41598-017-08265-7>
- 8 [40] Valeriani D and Poli R 2019 *PLOS ONE* **14** 1-17 URL
9 <https://doi.org/10.1371/journal.pone.0212935>
- 10 [41] Valeriani D and Matran-Fernandez A 2018 Past and future of multi-mind brain-computer interfaces
11 *Brain-Computer Interfaces Handbook: Technological and Theoretical Advances* ed Nam C S,
12 Nijholt A and Lotte F (CRC Press) chap 36, pp 685–700
- 13 [42] Pan S J and Yang Q 2009 *IEEE Transactions on knowledge and data engineering* **22** 1345–1359
- 14 [43] Waytowich N R, Lawhern V J, Bohannon A W, Ball K R and Lance B J 2016 *Frontiers in*
15 *neuroscience* **10** 430
- 16 [44] Ming Y, Ding W, Pelusi D, Wu D, Wang Y K, Prasad M and Lin C T 2019 *Applied Soft Computing*
17 **84** 105689
- 18 [45] Dai M, Zheng D, Liu S and Zhang P 2018 *Computational and mathematical methods in medicine*
19 **2018**
- 20 [46] Jayaram V, Alamgir M, Altun Y, Scholkopf B and Grosse-Wentrup M 2016 *IEEE Computational*
21 *Intelligence Magazine* **11** 20–31
- 22 [47] Rodrigues P L C, Jutten C and Congedo M 2018 *IEEE Transactions on Biomedical Engineering*
23 **66** 2390–2401
- 24 [48] Rolls E T, Grabenhorst F and Deco G 2010 *Neuroimage* **53** 694–706
- 25 [49] Valeriani D, Poli R and Cinel C 2015 A collaborative brain-computer interface for improving group
26 detection of visual targets in complex natural environments *2015 7th International IEEE/EMBS*
27 *Conference on Neural Engineering (NER)* (IEEE) pp 25–28
- 28 [50] Bhattacharyya S, Valeriani D, Cinel C, Citi L and Poli R 2019 Target detection in video feeds
29 with selected dyads and groups assisted by collaborative brain-computer interfaces *2019 9th*
30 *International IEEE/EMBS Conference on Neural Engineering (NER)* pp 159–162 ISSN 1948–
31 3546
- 32 [51] Bhattacharyya S, Valeriani D, Cinel C, Citi L and Poli R 2019 Collaborative brain-computer
33 interfaces to enhance group decisions in an outpost surveillance task *2019 41st Annual*
34 *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*
35 pp 3099–3102 ISSN 1557-170X
- 36 [52] Cui Y, Xu Y and Wu D 2019 *IEEE Transactions on Neural Systems and Rehabilitation Engineering*
37 **27** 2263–2273
- 38 [53] Wu D, Lance B J, Lawhern V J, Gordon S, Jung T and Lin C 2017 *IEEE Transactions on Neural*
39 *Systems and Rehabilitation Engineering* **25** 2157–2168
- 40 [54] Fernandez-Vargas J, Kita K and Yu W 2016 *Frontiers in Robotics and AI* **3** 50 ISSN 2296-9144
41 URL <http://journal.frontiersin.org/article/10.3389/frobt.2016.00050>
- 42 [55] Jervis B W, Ifeachor E C and Allen E M 1988 *Medical & Biological Engineering & Computing* **26**
43 2–12 ISSN 0140-0118 URL <http://link.springer.com/10.1007/BF02441820>
- 44 [56] Luce R D R D 1986 *Response times : their role in inferring elementary mental organization*
45 (Oxford University Press) ISBN 9780195036428
- 46 [57] Kiani R, Cortelli L and Shadlen M 2014 *Neuron* **84** 1329–1342 ISSN 0896-6273
- 47 [58] Fernandez-Vargas J, Tarvainen T V J, Kita K and Yu W 2017 *IEEE Access* **5** 23736–23750 ISSN
48 2169-3536 URL <http://ieeexplore.ieee.org/document/8081763/>
- 49 [59] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 *Journal of Machine*
50 *Learning Research* **15** 1929–1958 URL <http://jmlr.org/papers/v15/srivastava14a.html>
- 51 [60] Fleming S M, Huijgen J and Dolan R J 2012 *Journal of Neuroscience* **32** 6117–6125
- 52 [61] Fleming S M and Lau H C 2014 *Frontiers in Human Neuroscience* **8** 443 ISSN 1662-5161 URL
53 <http://journal.frontiersin.org/article/10.3389/fnhum.2014.00443/abstract>
- 54 [62] Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A and Yger F 2018 A
55 review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update
- 56
- 57
- 58
- 59
- 60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4

5 URL <https://doi.org/10.1088/1741-2552/aab2f2>

6 [63] 2019 Deep learning-based electroencephalography analysis: A systematic review URL
7 <https://doi.org/10.1088/1741-2552/ab260c>

8 [64] Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (EEG)
9 classification tasks: A review URL <https://doi.org/10.1088/1741-2552/ab0ab5>

10
11 **Additional material**
12
13

14 **Table 1.** Accuracy grouped by confidence level and task
15
16

Task	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Mean
PATROL1	0.56	0.60	0.71	0.73	0.78	0.74	0.78	0.85	0.88	0.96	0.97	0.82
PATROL2	0.43	0.57	0.64	0.68	0.74	0.77	0.77	0.81	0.93	0.97	0.97	0.81
PATROL3	0.26	0.49	0.56	0.56	0.68	0.63	0.77	0.77	0.83	0.91	0.96	0.80
SUSPECT-DETECTION	0.40	0.51	0.57	0.64	0.72	0.76	0.84	0.81	0.84	0.87	0.88	0.78
OUTPOST	0.29	0.56	0.73	0.74	0.80	0.82	0.91	0.93	0.94	0.89	0.94	0.87
ASSISTED-PATROL	0.57	0.59	0.63	0.68	0.75	0.79	0.81	0.87	0.95	0.97	0.98	0.86
REALISTIC-SEARCH1	0.74	0.64	0.73	0.76	0.77	0.84	0.81	0.86	0.91	0.93	0.98	0.81
REALISTIC-SEARCH2	0.50	0.66	0.75	0.78	0.81	0.72	0.63	0.70	0.78	0.80	1.00	0.74
Mean	0.54	0.59	0.67	0.70	0.76	0.76	0.81	0.85	0.90	0.92	0.96	0.82

43 **Table 2.** Response Time [s] grouped by confidence level and task
44

Task	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	Mean
PATROL1	1.4	1.2	1.1	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.8	1.0
PATROL2	1.0	1.3	1.3	1.2	1.1	1.2	1.1	1.2	1.0	0.95	0.95	1.1
PATROL3	1.1	1.2	1.2	1.2	1.2	1.1	1.1	1.1	1.0	0.95	0.95	1.1
SUSPECT-DETECTION	2.0	2.1	2.0	1.7	1.6	1.5	1.2	1.2	1.2	1.1	1.0	1.3
OUTPOST	2.3	2.1	1.8	1.8	1.9	1.9	1.8	1.7	1.6	1.4	1.2	1.6
ASSISTED-PATROL	1.1	1.1	1.1	1.2	1.2	1.3	1.3	1.4	1.3	1.2	1.1	1.2
REALISTIC-SEARCH1	1.6	1.5	1.4	1.6	1.4	1.3	1.3	1.2	1.2	1.0	1.0	1.3
REALISTIC-SEARCH2	1.4	1.6	1.4	1.3	1.5	1.8	1.5	1.1	1.0	1.0	1.0	1.4
Mean	1.3	1.4	1.3	1.3	1.3	1.4	1.3	1.3	1.2	1.1	1.0	1.2

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

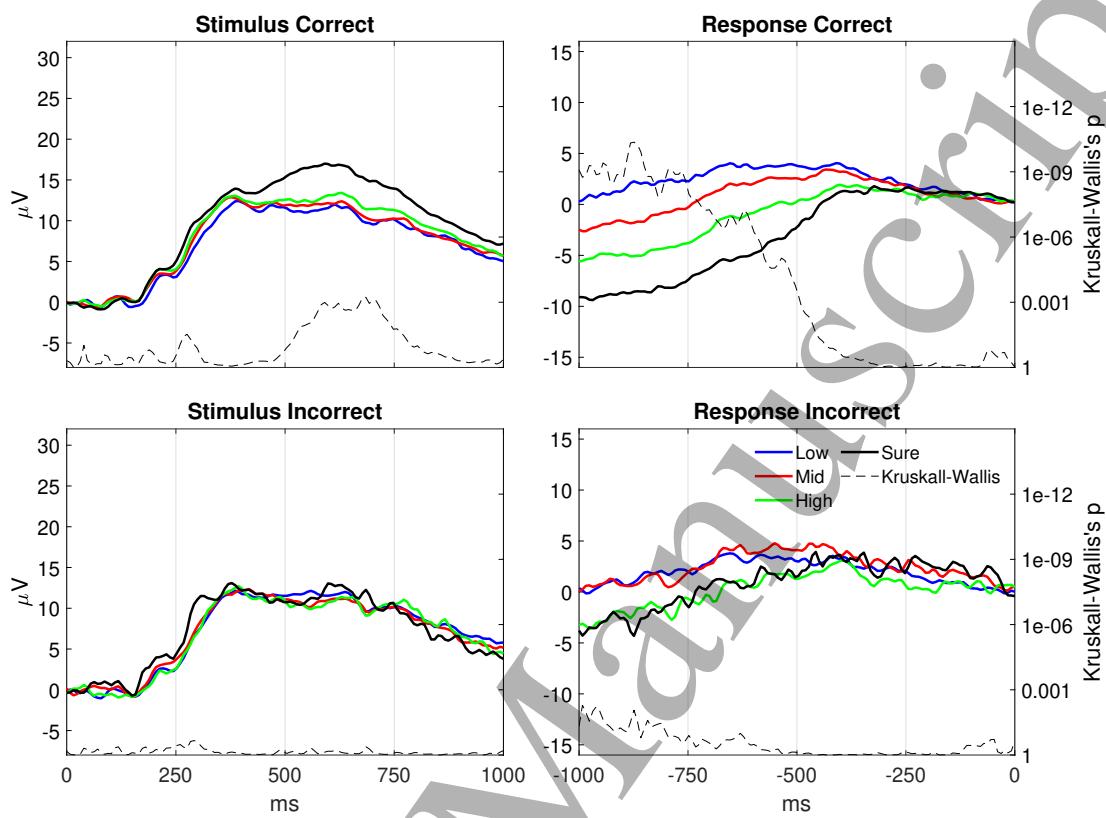
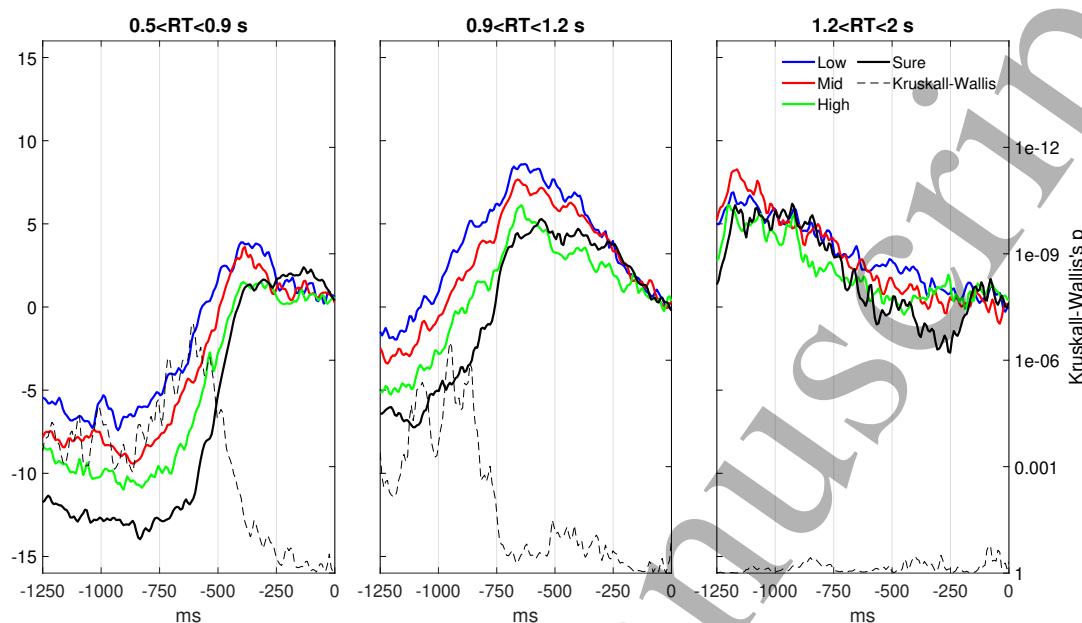


Figure 7. Stimulus (left) and Response (Right) locked grand average epochs for correct (above) and incorrect (below) trials.

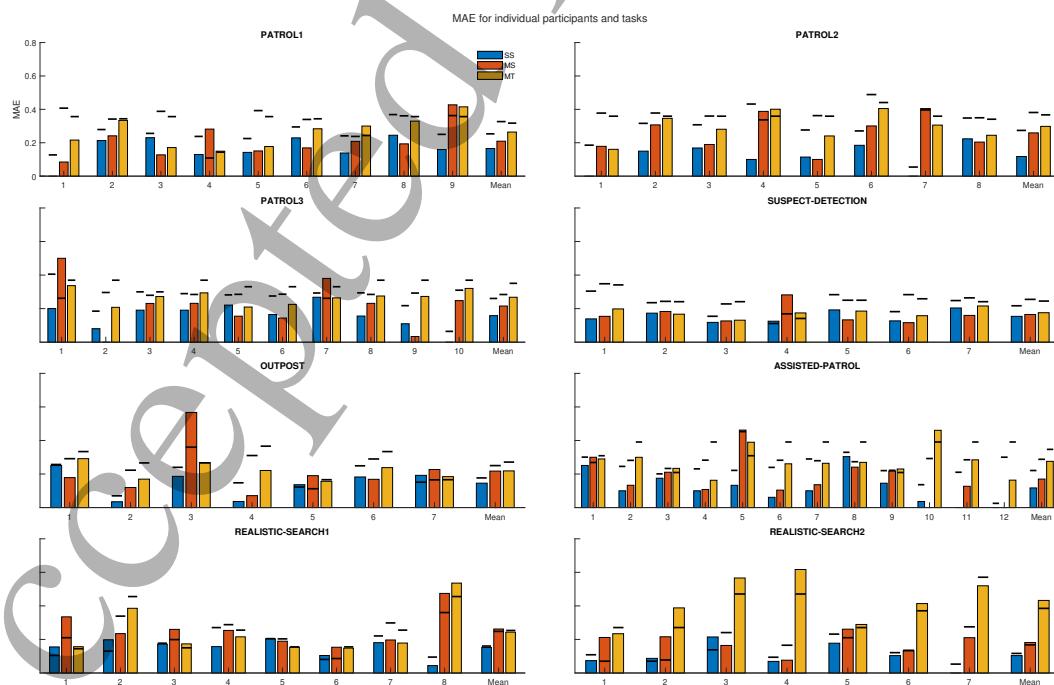
Table 3. Proportion of each confidence level reported grouped by task

Task	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
PATROL1	0.04	0.09	0.09	0.08	0.07	0.07	0.08	0.07	0.05	0.04	0.32
PATROL2	0.01	0.16	0.10	0.06	0.04	0.04	0.04	0.04	0.04	0.03	0.45
PATROL3	0.01	0.07	0.06	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.47
SUSPECT-DETECTION	0.00	0.03	0.06	0.09	0.09	0.11	0.10	0.10	0.12	0.08	0.20
OUTPOST	0.00	0.04	0.06	0.04	0.05	0.08	0.10	0.09	0.09	0.14	0.30
ASSISTED-PATROL	0.02	0.08	0.06	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.45
REALISTIC-SEARCH1	0.02	0.12	0.12	0.09	0.09	0.08	0.11	0.14	0.10	0.06	0.07
REALISTIC-SEARCH2	0.00	0.24	0.16	0.20	0.18	0.06	0.05	0.04	0.02	0.01	0.04
Mean	0.02	0.10	0.08	0.08	0.07	0.07	0.07	0.07	0.06	0.06	0.33

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3



26 **Figure 8.** Response locked grand average epochs grouped by RT.
27
28
29
30
31
32
33



58 **Figure 9.** MAE for individual participant and tasks. Bars indicate the result for the
59 ANN approaches, while the black horizontal lines indicate the baseline levels.
60

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

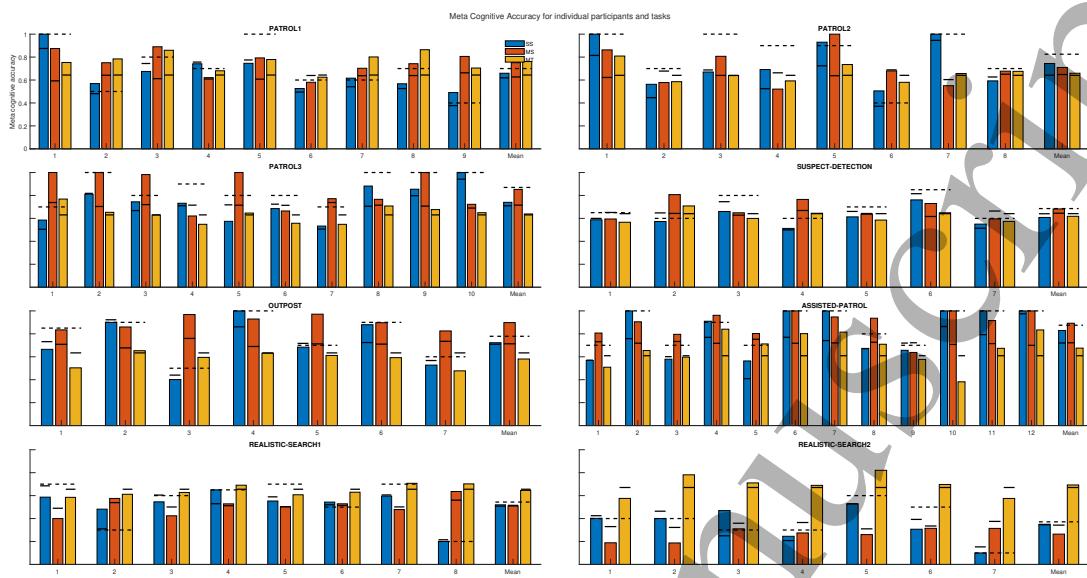


Figure 10. Metacognitive accuracy for individual participant and tasks. Bars indicate the result for the ANN approaches, while the black horizontal lines indicate the baseline levels. Dashed lines indicate value when the reported confidence was used (instead of the predicted one).

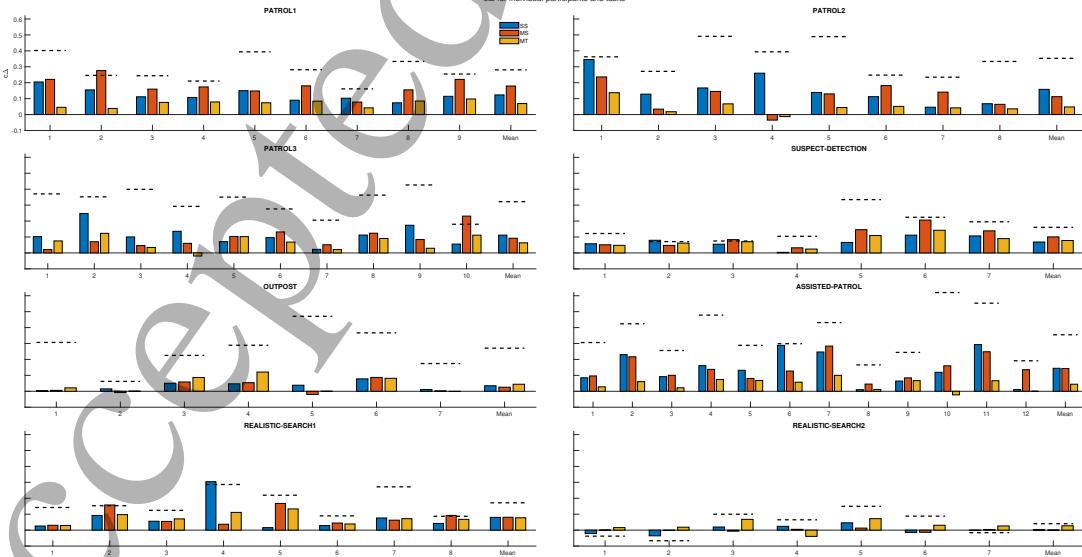


Figure 11. $c\Delta$ for individual participant and tasks. Bars indicate the result for the ANN approaches, while dashed lines indicate value when the reported confidence was used (instead of the predicted one). The baseline is not indicated because it's always 0.

1
2 *Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision m*
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

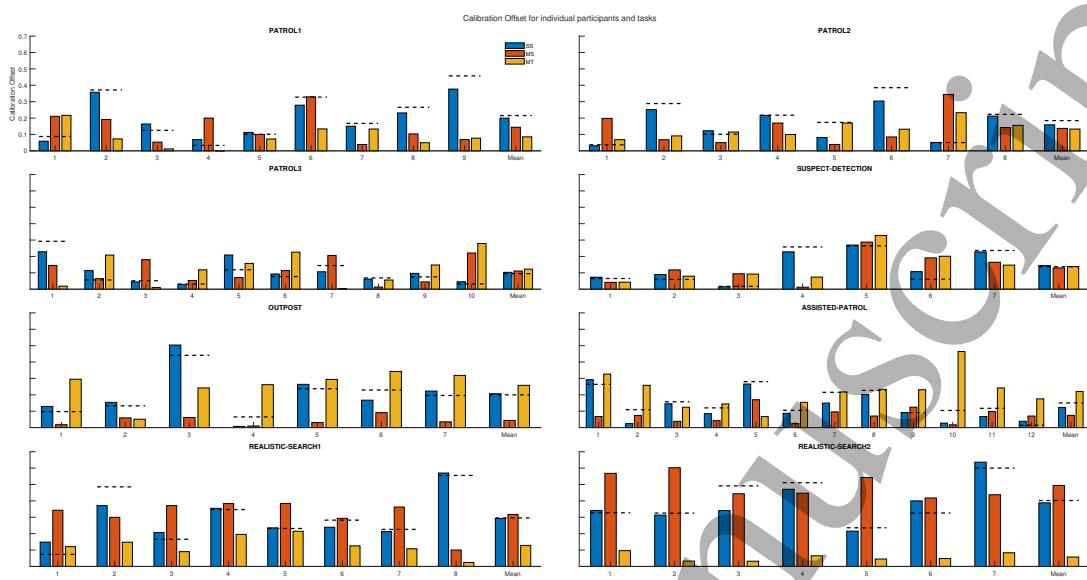


Figure 12. Calibration offset for individual participant and tasks. Bars indicate the result for the ANN approaches, while dashed lines indicate value when the reported confidence was used (instead of the predicted one).