Preparation and execution of voluntary action both contribute to awareness of intention

M. Schultze-Kraft, E. Parés-Pujolràs, K. Matić, P. Haggard, J.-D. Haynes

Proceedings of the Royal Society B

DOI: 10.1098/rspb.2019.2928

Electronic Supplementary Material

1. EEG-informed selection of participants

The readiness potential was the target brain signal that we aimed to use in order to manipulate prospective information about motor preparation. Therefore, we investigated whether such BCI-based manipulation was effective in each individual participant.

A qualitative assessment of EEG data from stage I used to train the classifier (Fig. S1A) shows that for most participants the signals look as expected, with EEG signals preceding self-paced movements in average displaying the typical negative trend of a readiness potential ("Move" class), while EEG signals preceding trial start cues do not show any particular trend ("Idle" class). However, while the RP is a potentially informative feature that the BCI may use for classification, it is not guaranteed a priori that the EEG features extracted by the classifier to separate the "Move" and "Idle" classes were based on the presence and absence of an RP over central channels.

A visual inspection of stage II data (Fig. S1B) suggests that for most participants *RP*+ cues were effectively preceded by an RP-like negativity, while *RP*- cues were not (with some conspicuous exceptions). Note that here we only consider *RP*+ and *RP*- cues that were elicited *before* any movement onset, thus excluding EEG data that would otherwise be contaminated with signals related to movement execution. In order to test whether we could rely on the BCI-triggered cues during stage II to discriminate *RP*+ from *RP*- activity in each individual participant, we performed the following analysis. Channel Cz was chosen for analysis because readiness potentials preceding foot movements are typically most distinct over that channel (Brunia et al., 1985). For each trial individually, we subtracted the average

EEG signal in the time interval -200ms to 0ms from the average EEG signal in the time interval -1200ms to -1000ms, with respect to the time of cue presentation. These values represent the relative change in amplitude in channel Cz during the 1.2 seconds before cue presentation. If the BCI relied on the readiness potential for classification, EEG signals over central channels preceding *RP*+ cues should be on average more negative than signals preceding *RP*- cues. To test this hypothesis, we ran a two-sample one-sided t-test, for each participant separately. The box plots in Fig. S1C show, for each participant and for *RP*+ and *RP*- cues individually, the distributions of amplitude changes, with participants ordered by the t-statistics of the t-test from largest (left) to smallest (right). For the first 16 participants the t-test showed that signals preceding *RP*+ cues became significantly more negative during the 1.2 sec interval than signals preceding *RP*- cues. For the remaining 7 participants this was not the case, suggesting that the classifier did not made interruptions based on the presence or absence of RP-like events in the EEG. Consequently, these participants are excluded from all subsequent analyses. Fig. S1D shows individual and grand average EEG signals preceding *RP*+ and *RP*- cues of the 16 participants selected for the final sample.

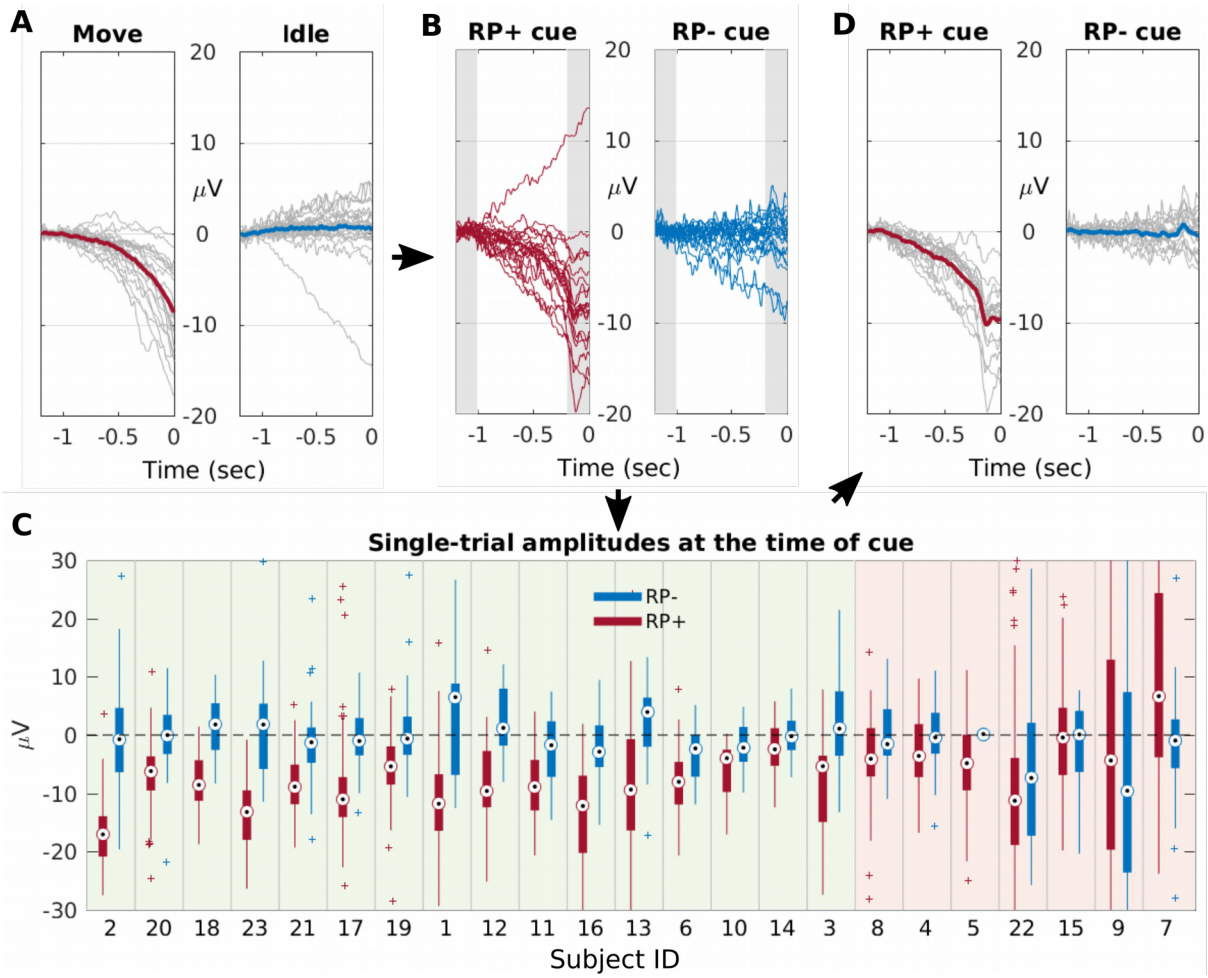

Fig. S1. Selection of participants based on EEG signals of channel Cz. (**A**) Event-related potentials (ERPs) of EEG signals recorded during stage I, time-locked to self-paced EMG onsets (left, "Move") and time-locked to trial start cues (right, "Idle"). ERPs are baseline corrected in the interval -1200 to -1000 ms, and shown for individual participants (grey) and as grand average (colored). (**B**) ERPs of EEG signals recorded in stage II, time-locked to *RP*+ cues (left) and time-locked to *RP*- cues (right). ERPs are baseline corrected in the interval -1200 to -1000 ms and shown for individual participants. (**C**) For each participant (ID on x-axis) and for *RP*+ and *RP*- cues individually (color coded), the box plots show the distribution of EEG signal amplitude changes between the time interval -1200 to -1000 ms and -200 to 0 ms with respect to cue onset (indicated by gray areas in panel B). Participants are ordered in ascending order by the t-statistic of a two-sample one-sided t-test that tests whether the mean change in *RP*+ trials was more negative than in *RP*- trials. Participants for which $p<.01$ are highlighted in green, otherwise in red. (**D**) As in B, but only for the selected N=16 participants (gray) and the corresponding grand average (colored).

2. Setup of Real-time BCI

For the BCI predictor used in stage II, a linear classifier was trained on EEG data from the 100 pedal presses recorded during stage I.

2.1. EMG onset detection

For each trial, we assessed the movement onset. For higher temporal precision, we defined the onset of the movement based on the EMG rather than based on the final completion of the movement with the pedal press. To obtain EMG onset we high-pass filtered the EMG signal at 20 Hz. Then the standard deviation of the signal during the first 1000 ms after each trial start cue was determined as an *idle* baseline. For each trial individually, the standard deviation of subsequent, overlapping 50 ms windows was computed and EMG onset identified as the end of the first 50 ms window where the standard deviation exceeded idle baseline by a factor of 3.5.

2.2. Class specification

Based on these movement onsets, two periods were defined as *move* and *idle* for the training of the classifier. The *move* periods were 1200 ms long segments preceding EMG onset, while the *idle* periods were 1200 ms long segments preceding the trial start cue (i.e. the onset of the traffic light).

2.3. Feature extraction and classifier training (Stage I)

EEG data from those segments were baseline corrected to the mean signal in the time interval between -50 and 0 ms locked to EMG onset or trial start cue, respectively. These were then averaged over time windows defined by the time points -1200, -900, -650, -450, -300, -200, -100 and -50 ms locked to EMG onset or trial start cue, respectively. The choice of the baseline correction interval being locked to the end of the EEG segment (as opposed to the traditional choice of being locked to the beginning of the segment) and the choice of unequal time intervals were both based on a piloting analysis on previous data (Schultze-Kraft et al., 2016) that showed improved classification accuracy with these parameters. The resulting values were concatenated and used as features to train a regularized Linear Discriminant Analysis (LDA) classifier with automatic shrinkage (Blankertz et al., 2011). Classification accuracy (obtained from a 10-fold cross-validation on stage I data) averaged 81.8% (SEM = 2.1%) across participants. In Table S1 we report classification accuracies of single

participants. Fig. S2 shows, for channel Cz exemplarily, the single-trial EEG waveforms of the *move* periods used to train the classifier.

Participant	1	2	3	6	10	11	12	13
Accuracy	**86.9%**	**91.4%**	**89.6%**	**88.8%**	**82.0%**	**91.7%**	**75.5%**	**78.1%**
Participant	14	16	17	18	19	20	21	23
Accuracy	**71.1%**	**86.1%**	**86.7%**	**81.3%**	**79.9%**	**69.0%**	**88.4%**	**62.9%**

Tab. S1. Cross-validated classification accuracies of single participants. Classification accuracies were computed on the 100 trials from stage 1 in a leave-one-out cross-validation. Mean accuracy was 81.8% (SEM=2.1).

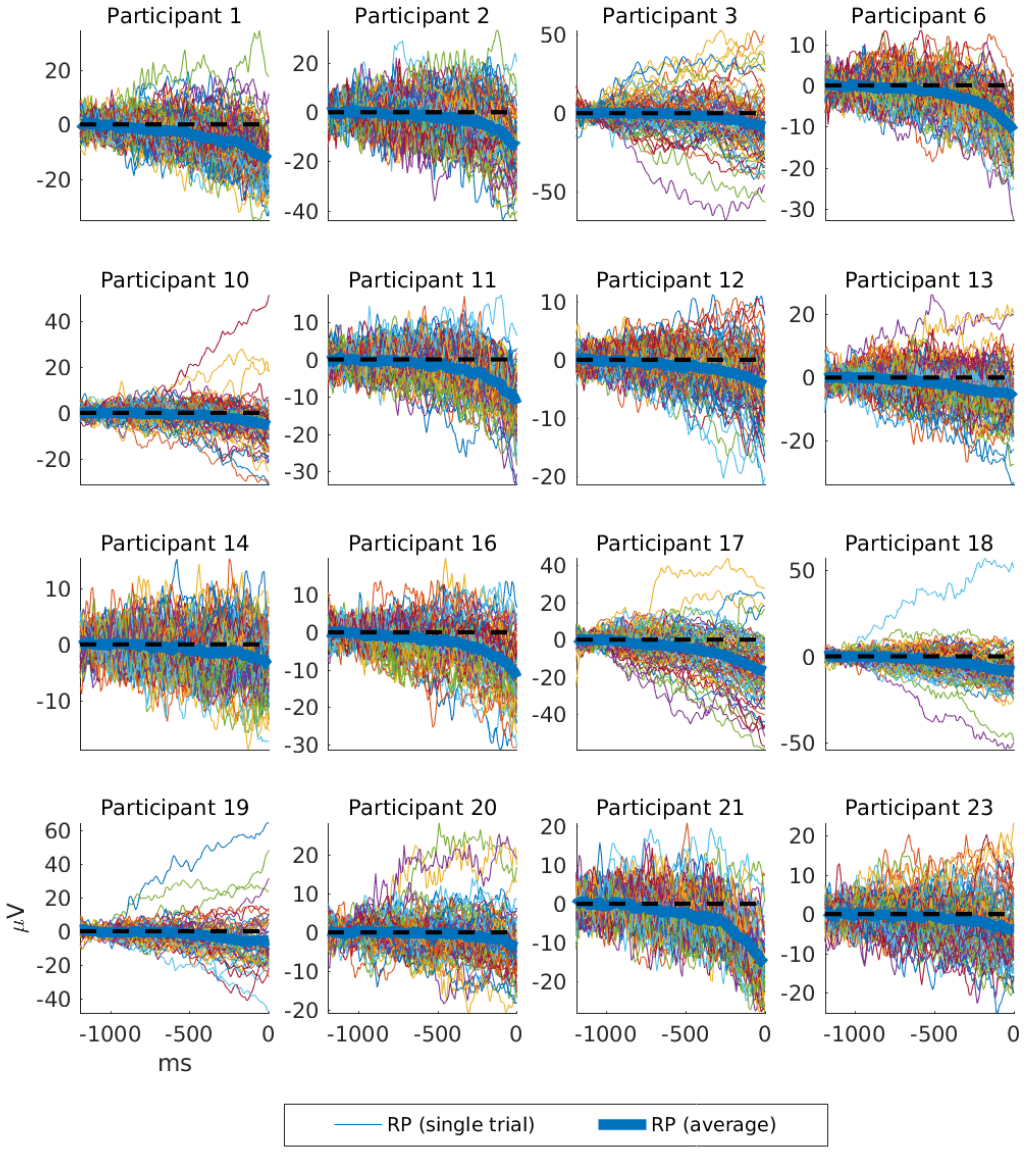

Fig. S2. Single-trial EEG waveforms of RPs in stage I. Panels show, for each participant individually, single-trial EEG waveforms in channel Cz recorded in stage I. These are time-locked to movement onset and thus constitute single-trial RPs. Baseline corrected in the interval [-1200 -1000] ms. Average is shown as a thick blue trace. Axes are as specified in the lower left panel.

2.4. Real-time application of classifier (Stage II)

During stage II, the so-trained classifier was used to monitor the ongoing EEG in real-time. Therefore, every 10 ms a feature vector was constructed from the immediately preceding 1200 ms of EEG data, as outlined above, and used as input to the classifier, generating a classifier output value every 10 ms. This output variable was a continuous signal that probabilistically classified the current EEG segment either to the *idle* or to the *move* class.

2.5. Threshold setting

A classifier threshold was set for each participant individually. Because the classifier output signal was likely to mirror the stochastic nature of the EEG, a conservative threshold was defined in order to avoid many cues to be prematurely triggered by noise. For this, we trained the classifier on 99 trials from stage I and applied it to each consecutive and overlapping 1200 ms feature window in the left out trial, thereby mimicking the real-time application during stage II. This was done for each of the 100 trials in a leave-one-out crossvalidation scheme. For each of these continuous classifier output vectors the time of first threshold crossing after trial start was computed. Let us refer to the time of first threshold crossing in a trial as a "prediction" event. Now, we define predictions occurring somewhere between trials start and up to 600 ms before movement onset as false alarms (FA), predictions occurring between 600 ms before movement onset and the time of movement onset as Hits, and predictions occurring after movement onset or not occurring at all as Misses. From this the F-measure (Powers, 2011) $F_\beta(\theta)=\frac{(1+\beta^2)\,Hit(\theta)}{(1+\beta^2)\,Hit(\theta)+\beta^2\,Miss(\theta)+FA(\theta)}$ was computed for different threshold values θ. The largest F thus corresponds to the threshold θ were the Hit rate is maximal, while at the same time the FA and Miss rates are minimal. Moreover, by choosing β=0.5, we aimed at giving the minimization of FAs more weight than minimizing Miss rate. We prioritized minimizing the number of false alarms, at the cost of potentially

missing some actions. The resulting F-values were smoothed and the threshold with the highest F-value chosen.

Fig. S3 shows single-trial EEG waveforms time-locked to *RP+* cues in (stage II) that were elicited when the specified threshold was reached before participants initiated a movement. In Fig. S4 we compare the distribution of amplitudes of the waveforms shown in Fig. S2 and S3, respectively.

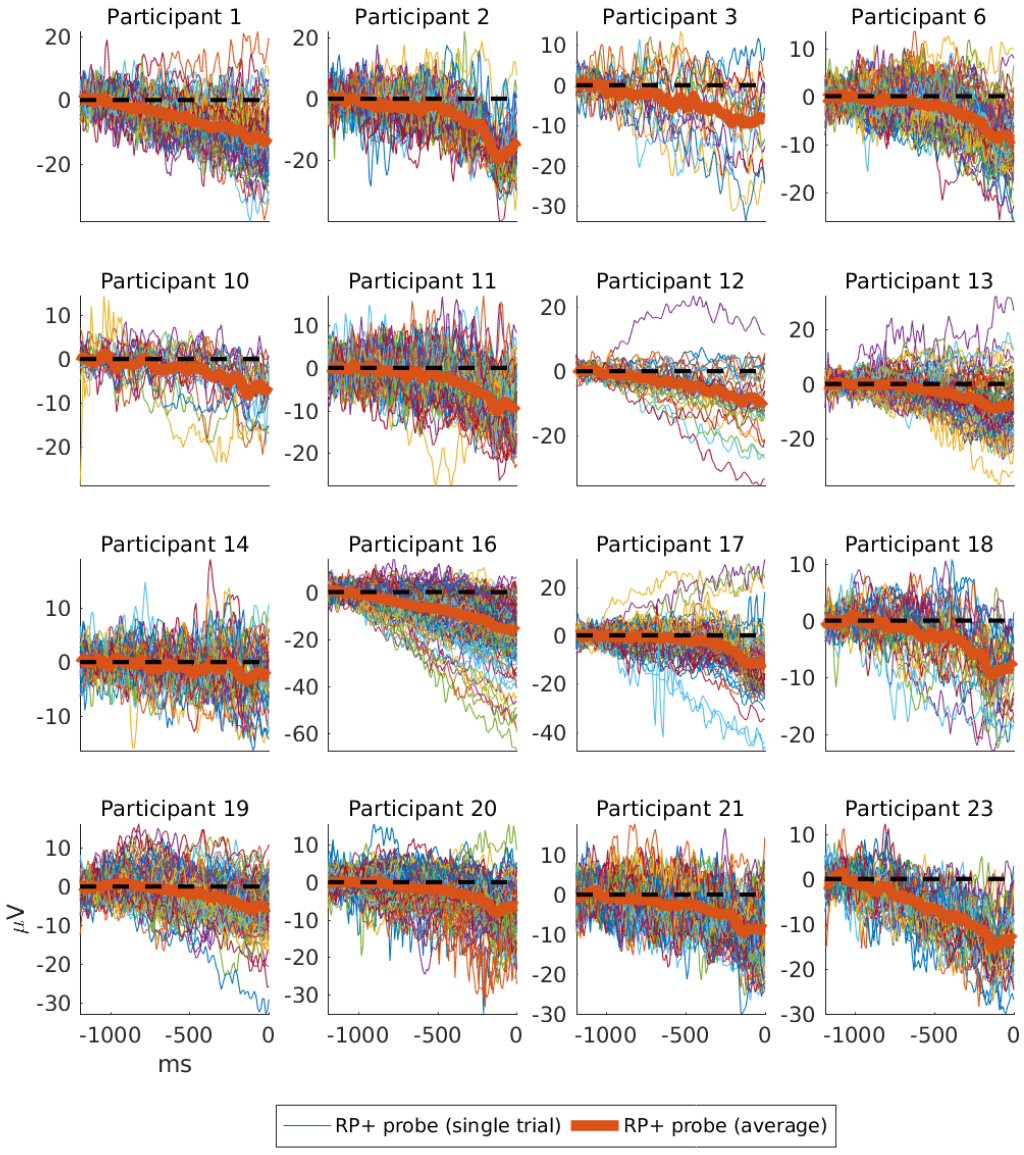

Fig. S3. Single-trial EEG waveforms of *RP+* cues in stage II. As in Fig. S2, but here EEG waveforms are time-locked to *RP+* cues elicited by the BCI during stage II. Average is shown in thick red traces. Axes are as specified in the lower left panel.

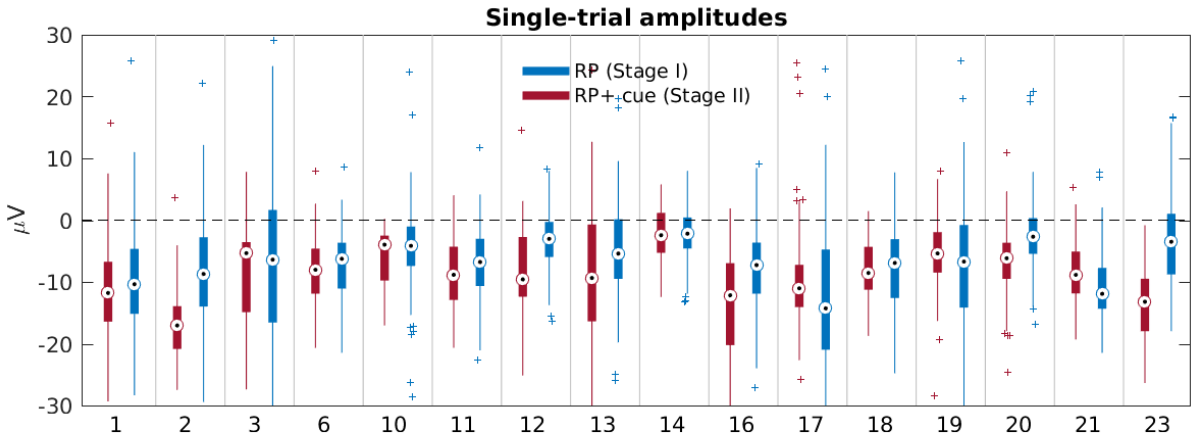

Fig. S4. Distribution of single-trial amplitudes. For each participant (ID on x-axis), the box plots show the distribution of EEG signal amplitude changes between the time interval -1200 to -1000 ms and -200 to 0 ms, with respect to movement onsets in stage I (blue), and with respect to *RP+* cues (stage II). The former thus represent single-trial amplitudes of the RP used to train the BCI.

3. Model selection procedure and statistical details

As described in the Statistical Analysis section of the methods, we used linear mixed-effects models to test the effects of our explanatory variables on the probability of participants reporting awareness. To select the model that best explained our observed results, we followed the random effect selection procedure suggested in (Matuschek at al. 2017).

In all models, a random intercept was included to account for the variability in the dependent variable across participants. Further, we included those random effects that significantly improved the model fit. To determine the optimal random effects structure, we fit a baseline model which included all explanatory variables and all possible interactions as fixed terms. We then iteratively compared this baseline models against models with one additional random slope using a chi-squared test. If the inclusion of a random slope significantly improved the model fit, the random slope was included in the final model. This approach has been suggested as a better option than including random slopes for all fixed effects, as it decreases the probability of Type II errors while maintaining the same power against type I errors, and has previously been used in the literature (e.g. Steinemann et al., 2018). All models were fit using the *glmer* function in the homonymous R package (Bates et al., 2015).

Tables S2 and S3 provide the detailed results of the random effect selection procedure for both main analyses and the final inference statistics reported in the main text.

Supplementary Table 2: model 1 random effects selection			
Test individual random effects	**Baseline model:** yes ~ 1 + RP + Action + RP:Action + (1+|sub)		
	X^2	**DF**	**p-value**
yes ~ 1 + RP + Action + RP:Action + (1+RP|sub)	10.939	2	0.0042 **
yes ~ 1 + RP + Action + RP:Action + (1+Action|sub)	17.07	2	<0.001***

Tab. S2. Model 1 selection steps and statistical results of model comparison. Random slopes for both RP and Action significantly improved the fit of the baseline model and were therefore included in the model.

Supplementary Table 3: model 2 random effects selection			
Test individual random effects	**Baseline model:** yes ~ 1 + RP + RT + RP:RT + (1+|sub)		
	X^2	**DF**	**p-value**
yes ~ 1 + RP + RT + RP:RT + (1+RP|sub)	0.2518	2	0.881
yes ~ 1 + RP + RT + RP:RT + (1+RT|sub)	7.1686	2	0.027*

Tab. S3. Model 2 random effect selection steps and statistical results of model comparison. Only the RT random slope significantly improved the fit of the baseline model and was therefore the only random effect included in the model.

4. Data description and trial selection procedure

The number of trials in which participants were presented a cue, as well as the exact times when cues were presented, could not be precisely experimentally controlled. In case of *RP+* trials, this is because the BCI was calibrated so as to elicit cues preferably during the interval just before a movement, based on the detection of a readiness potential. However, the detection of transient events in the EEG in real-time by means of an asynchronous BCI is only possible with a limited accuracy, bound by the noisy nature of EEG signals. Further, to test our hypothesis we required participants to correctly follow the instructions

Fig. S5 illustrates the types of trials that occurred during the task and highlights the ones that were included for analysis.

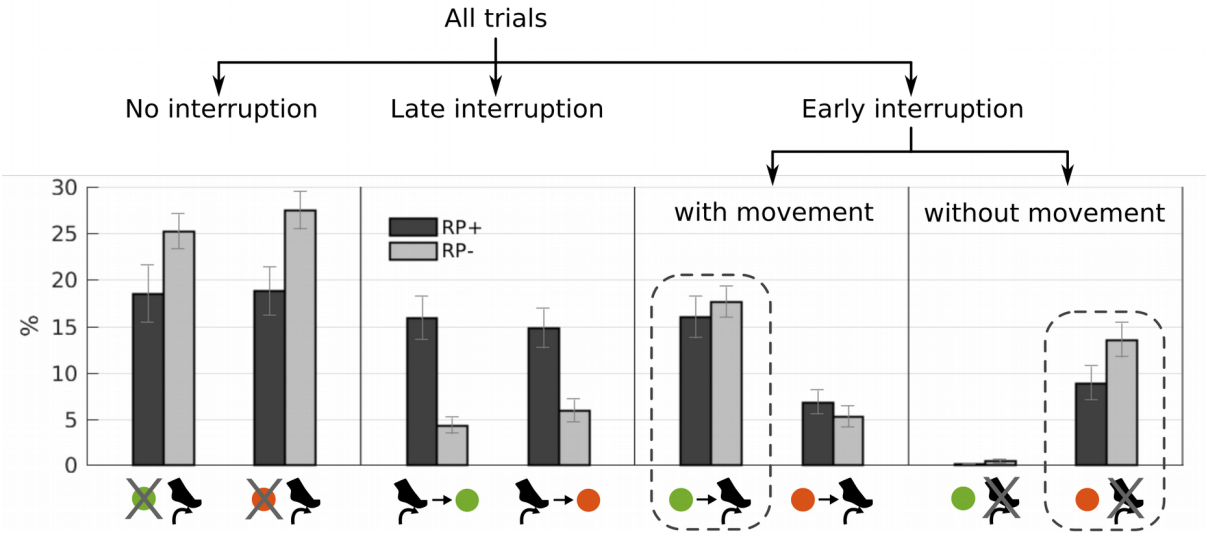

Fig. S5. Types of observed trials and selection procedure. Bar graphs represent the grand-averaged percentage (+ SEM) of trials within each category, for *Go* (green cues) and *No-Go* (red cues) trials, and in the *RP+* (dark grey) and *RP-* condition (light grey). Percentages are calculated in *RP+* and *RP-* conditions separately, i.e. dark grey and light grey bars sum up to 100, respectively. The pictograms below the bar graphs indicate the temporal relation between cue presentation and movement onset. *No interruption* trials: In some trials, participants executed a movement and no cue was presented at all. In these trials, no awareness report was collected and no further analysis was conducted. *Late interruption* trials: In some trials, cues came "too late", shortly after participants had already started moving. All these trials were discarded from further analysis. *Early interruption* trials: Cues were shown before any EMG onset was detected. In the *Go* condition, only trials where participants moved after the green cue presentation were included for analysis (left dashed

box). In the *No-Go* condition, only trials where no EMG onset was detected after the red cue presentation were included for analysis (right dashed box).

4.1. Characterization of trial types

In some cases, participants pressed the pedal *without* a cue being elicited (*No interruption* trials). In the *RP+* condition, these represent instances where the BCI failed to detect a readiness potential (*RP+/Go*: M = 18.5%, SEM = 3.1%; *RP+/No-Go:* M = 18.8%, SEM = 2.6%). In *RP-* trials, these represent instances where participants pressed the pedal before the random predetermined time of the cue (*RP-/Go:* M = 25.3%, SEM = 1.9%; *RP-/No-Go:* M = 27.5%, SEM = 2.0%). In all these cases, since no cue was presented, no awareness report was collected. Thus these trials are excluded from further analysis.

In another subset of trials, a cue was presented *after* EMG onset (*Late interruption* trials). In some *RP+* trials, a readiness potential was presumably correctly detected by the BCI, but a cue was presented after participants had already started moving (*RP+/Go:* M = 15.9%, SEM = 2.3; *RP+/No-Go:* M = 14.8%, SEM = 2.1). In turn, the *RP-* trials where a cue was presented after participants' movement reflect rare instances where the predetermined probing time by chance coincided with the self-paced time of movement (*RP-/Go*: M = 4.3%, SEM = 0.9; *RP-/No-Go:* M = 5.9%, SEM = 1.2). For our purposes, these cues came too late and the corresponding awareness reports are thus excluded from further analysis.

In another subset of trials, the cue was presented *before* EMG onset. In the *Go* condition (*Early interruption* trials *with movement*), these trials fulfil our prerequisite that *Go* cues must be followed by a movement, and thus the corresponding awareness reports are used in the main analysis (*RP+/Go:* M = 16.0%, SEM = 2.2; *RP-/Go:* M = 17.7%, SEM = 1.7). In the *No-Go* condition, participants sometimes initiated a movement after a cue was presented (*RP+/No-Go:* M = 6.9%, SEM = 1.3; *RP-/No-Go:* M = 5.3%, SEM = 1.1). Although they were often able to abort a movement before fully pressing the pedal in some of these trials, the very initiation of a movement - even though aborted - might suffice for participants to reconstruct an awareness of intention in the awareness probes that followed those cues. Thus, these trials were excluded from further analysis.

Finally, in some trials a cue was elicited before any EMG onset but no movement was produced after it (*Early interruption* trials *without movement*). In the *Go* condition, these very rare occurrences reflect trials where participants failed to respond with a pedal press to a

green cue (*RP+/Go*: M = 0.15%, SEM = 0.06; *RP-/Go*: M = 0.4%, SEM = 0.2). In contrast, as expected, in the *No-Go* condition this occurred more frequently (*RP+/No-Go:* M = 8.9%, SEM = 1.8; *RP-/No-Go:* M = 13.6%, SEM = 1.8). In these trials, participants successfully followed the instruction to withhold any movement after a red cue. Because they fulfil our prerequisite that *No-Go* cues must not be followed by a movement, the corresponding awareness reports are used in the main analysis.

4.2. Time relation between movements and cues

A closer look into the distributions of time differences between EMG onsets and cues provides further insight into the way in which our experimental design resulted in the observed proportions of trials (Fig. S6).

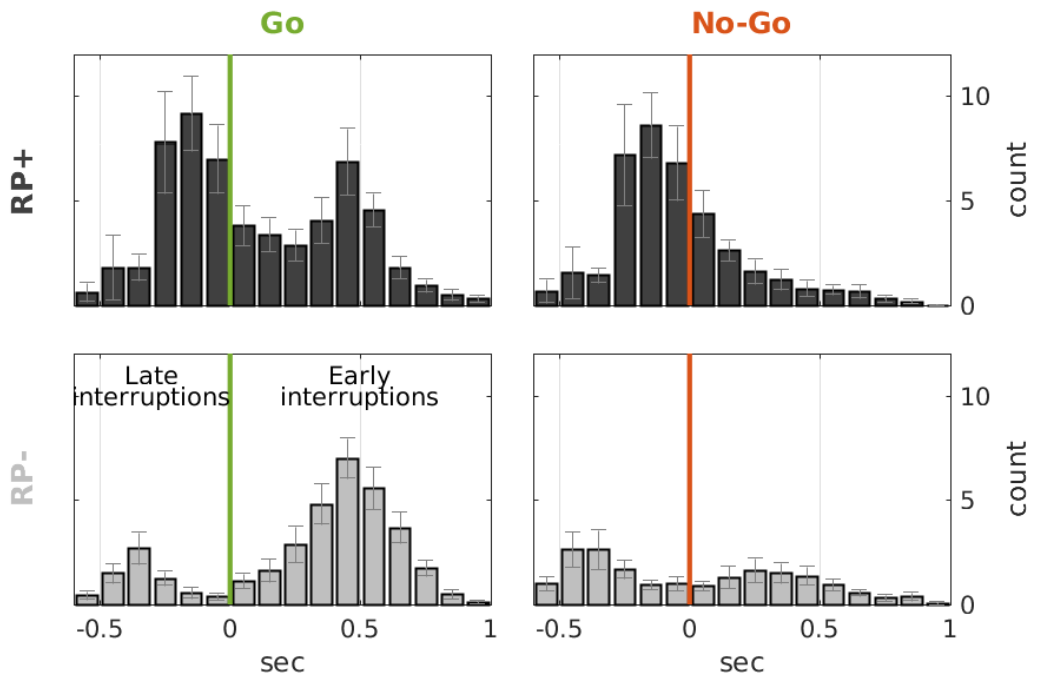

Fig. S6. Reaction times in trials with cues and movements. Shown are the time histograms of EMG onset time with respect to the time of cue presentation in *Go* (left) and *No-Go* (right) trials, in the *RP+* (top) and *RP-* condition (bottom). Negative times correspond to EMG onsets in *Late interruption* trials, in which the cue was presented after participants started moving. Positive times correspond to the distribution of EMG onsets in *Early interruption* trials (i.e. classic reaction times), where a cue was presented and a movement was initiated shortly afterwards. Bars and antennas show the grand averages and SEMs of trial counts in 100 ms bins, respectively.

Go and *No-Go* signals were often triggered *after* participants had started moving in the *RP+* conditions, while this was rarely the case in *RP-* condition. In the *RP+* condition, these *Late interruption* trials correspond to the distribution centered before cue presentation. These are instances of motor preparation states that were successfully detected by the BCI, but too late. Trials falling on the right tail of this distribution were instead instances where motor preparation was successfully interrupted early by the BCI. These trials can be interpreted as interruptions after the point of no return (Schultze-Kraft et al., 2016). That is, trials in which participants would have moved *anyway* if a cue had not been presented. In fact, in a number of *No-Go* trials participants failed to inhibit a movement and an EMG onset was detected after the red cue. In turn, in the *Go* condition, the effect of these intercepted self-paced actions is visible in the higher count of trials with very fast responses (RT < 200 ms) in the *RP+/Go* condition compared to the *RP-/Go* condition. In sum, in the *RP+/Go* condition, very fast trials (<200 ms) include both self-paced movements that happened to occur just after the green *Go* signal (right tail of the *Late interruption* distribution, Fig. S5), and also reactions to the *Go* signal (left tail of the *Early interruption* distribution). In turn, in the *RP-/Go* condition, movements produced very fast after the cue presentation were only reactions.

We checked that these very fast responses in the left tail of the *Early interruption* distribution could physiologically be fast reactions rather than self-paced actions that the classifier did not predict, by looking at the RT distribution during a simple cued reaction time task. These RTs were recorded in a final stage of the experiment, where no self-paced actions were being performed and participants were only reacting to *Go* cues presented at random times (Fig. S7). Here, we also observed some very fast reaction times (<200 ms) comparable to the ones found in the *RP-* condition.

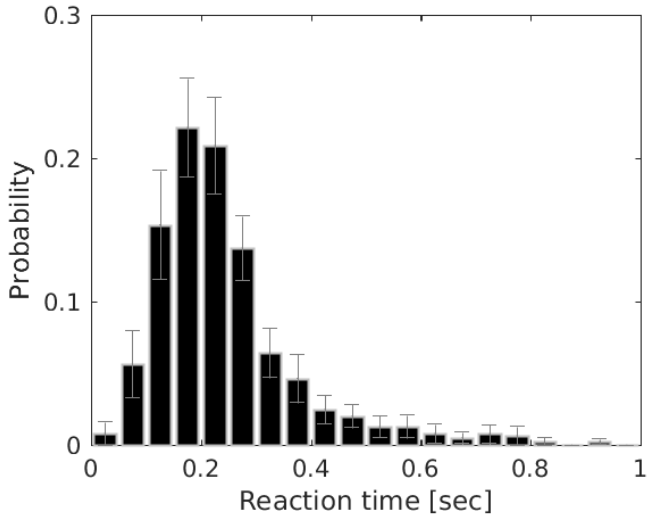

Fig. S7. Distribution of reaction times in a simple cued reaction task. The histogram shows, in discrete 50 ms bins, the probability (± SEM) of observing EMG onsets after presentation of the Go cue.

5. Control of instruction effects in model 1

To check that the observed retrospective effects were not merely driven by the instructions rather than the action execution, we ran a mixed-model predicting the probability of reporting awareness based on the presence or absence of the RP (*RP*+/*RP*-) and on whether participants moved (EMG+) or did not move (EMG-) after the *No-Go* cue was presented (Fig. S8). A similar analysis was not performed in the *Go* condition because participants extremely rarely failed to execute an action following the instruction to press the pedal after the green light. Intention reports were significantly more likely after a movement than in its absence within the *No-Go* condition (X^2 = 200.23, p < 0.001). This suggests that participants' reports were not merely driven by the instruction to move or not move.

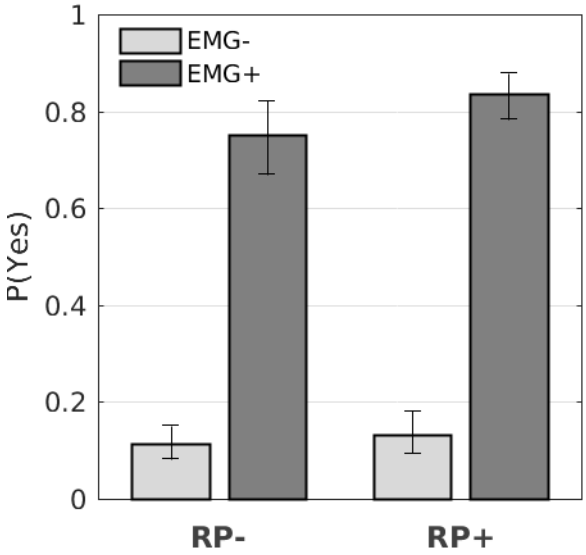

Fig. S8. Retrospective effects are not driven by instruction. Bars represent responses in trials where a *No-Go* cue was not followed by an EMG onset (EMG-, included in main analysis) or was followed by an EMG onset (EMG+, excluded from the main analysis), for *RP*- (left) and *RP*+ (right) conditions separately. Participants were significantly more likely to report an intention to move after an EMG onset than in the absence of it within the *No-Go* condition. Bars and antennas show probability estimates and 95% confidence intervals, respectively, calculated by pooling the responses of the corresponding subset of trials across all participants.

6. Testing for effect of time of cue in models 1 and 2

Supplementary Table 4: model comparison including cue presentation time			
(A) Model comparison	**Original model** P(yes) ~ 1 + RP + Action + RP:Action + (1+Action + RP|sub)		
	X^2	**DF**	**p-value**
Control model yes ~ 1 + RP + Action + CT + RP:CT + RP:Action + RP:Action + RP:Action:CT + (1+Action + RP|sub)	3.95	4	0.4127
(B) Control model output	**X^2**	**DF**	**p-value**
RP	9.53	1	0.002**
Action	19.89	1	<0.001***
CT	2.63	1	0.104
RP:Action	2.10	1	0.1467
RP:CT	1.31	1	0.2507
Action:CT	0.09	1	0.7546
RP:Action:CT	0.21	1	0.6402

Tab. S4. To control whether the effects observed in model 1 could be accounted for by the time at which the *RP+*/*RP-* cue were presented, we included the Cue Time (CT) as a fixed effect in the model, together with its interaction with RP and Action. The inclusion of this variable did not significantly improve the model fit (A), and the statistical significance of the Action and RP effects remained unchanged (B).

Supplementary Table 5: model comparison including cue presentation time			
(A) Model comparison	**Original model** P(yes) ~ 1 + RP + RT + RP:RT + (1+ RT |sub)		
	X^2	**DF**	**p-value**
Control model yes ~ 1 + RP + RT + CT + RP:CT + RP:RT + RT:CT + RP:RT:CT + (1+RT |sub)	6.09	4	0.1919
(B) Control model output	**X^2**	**DF**	**p-value**
RP	1.08	1	0.29
RT	65.29	1	<0.001***
CT	1.44	1	0.23
RP:RT	8.33	1	0.003**
RP:CT	1.31	1	0.25
RT:CT	0.95	1	0.32
RP:RT:CT	1.88	1	0.1688

Tab. S5. To control whether the effects observed in model 2 could be accounted for by the time at which the *RP+*/*RP-* cue were presented, we included the Cue Time (CT) as a fixed effect in the model, together with its interaction with RP and RT. The inclusion of this variable did not significantly improve the model fit (A), and the statistical significance of the RT and RP effects remained unchanged (B).

7. Single subject reaction times and response probabilities in *Go* trials

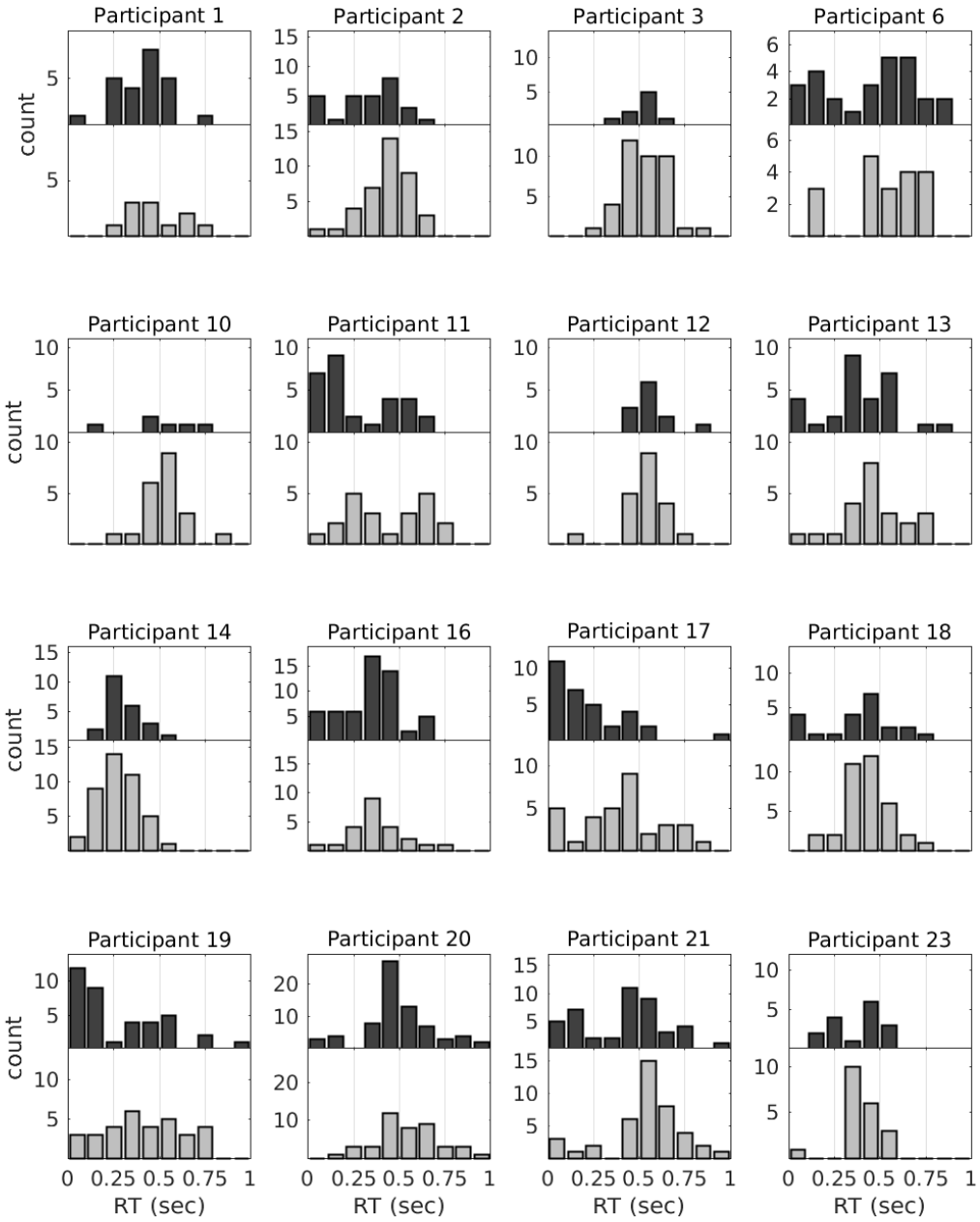

Fig. S9. Reaction time distribution in *Go* trials for each individual participant in the *RP*+ (black) and *RP*- (gray) conditions.

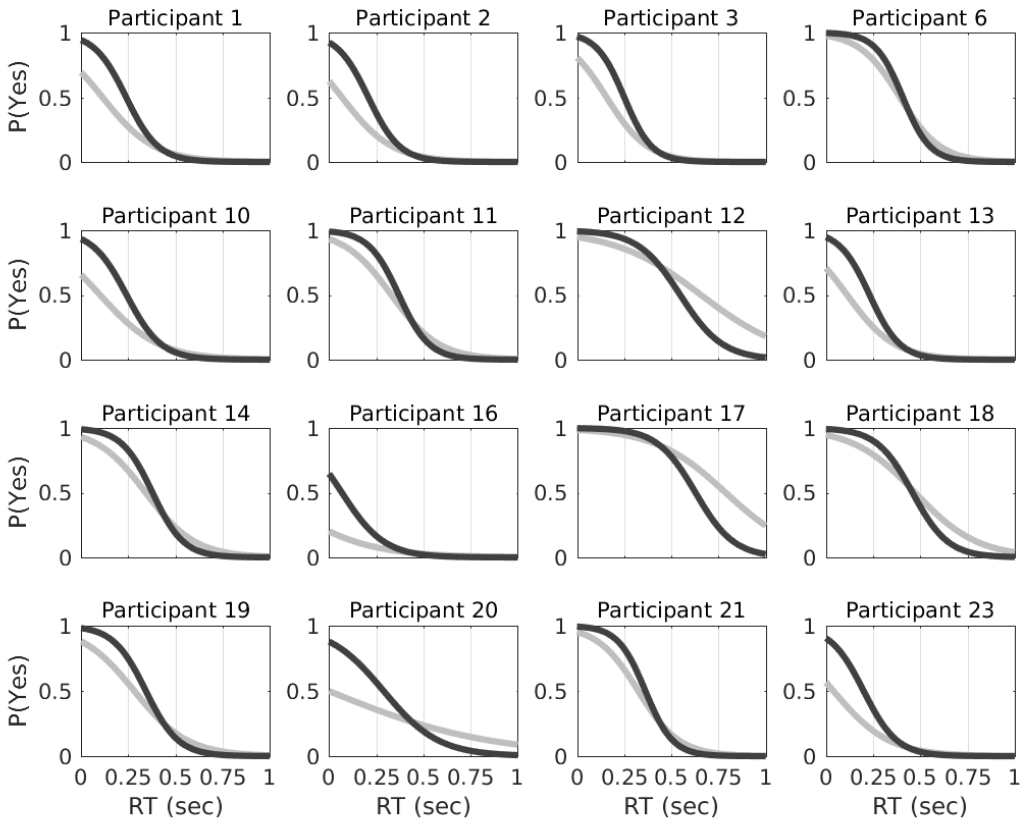

Fig. S10. Probability of responding 'yes' in *Go* trials as predicted from a regression model fitted to each individual participant in the *RP+* (black) and *RP-* (gray) conditions.

References

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015 Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**, 1-48.

Blankertz, B., Lemm, S., Treder, M. S., Haufe, S. & Müller, K.-R. 2011 Single-trial analysis and classification of ERP components - a tutorial. *NeuroImage* 56, 814-825.

Brunia, C. H., Voorn F. J. & Berger, M. P. 1985 Movement related slow potentials. II. A contrast between finger and foot movements in left-handed subjects. *Electroen Clin Neuro* **60**, 135-145.

Matuschek, H, Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. 2017 Balancing Type I error and power in linear mixed models. *J Mem Lang* **94**, 305-315.

Powers, D. 2011 Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. *J Mach Learn Technol* **2**, 37-63.

Steinemann, N. A., O'Connell, R. G. & Kelly, S. P. 2018 Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nat Commun* **9**, 3627.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., Blankertz, B. & Haynes, J.-D. 2016 The point of no return in vetoing self-initiated movements. *Proc Natl Acad Sci U S A* **113**, 1080-1085.