

OPEN

Decomposing Simon task BOLD activation using a drift-diffusion model framework

James R. McIntosh ^{1*} & Paul Sajda ^{1,2}

The Simon effect is observed in spatial conflict tasks where the response time of subjects is increased if stimuli are presented in a lateralized manner so that they are incongruous with the response information that they represent symbolically. Previous studies have used fMRI to investigate this phenomenon, and while some have been driven by considerations of an underlying model, none have attempted to directly tie model and BOLD response together. It is likely that this is due to Simon models having been predominantly descriptive of the phenomenon rather than capturing the full spectrum of behavior at the level of individual subjects. Sequential sampling models (SSM) which capture full response distributions for correct and incorrect responses have recently been extended to capture conflict tasks. In this study we use our freely available framework for fitting and comparing non-standard SSMs to fit the Simon effect SSM (SE-SSM) to behavioral data. This model extension includes specific estimates of automatic response bias and a conflict counteraction parameter to individual subject behavioral data. We apply this approach in order to investigate whether our task specific model parameters have a correlate in BOLD response. Under the assumption that the SE-SSM reflects aspects of neural processing in this task, we go on to examine the BOLD correlates with the within trial expected decision-variable. We find that the SE-SSM captures the behavioral data and that our two conflict specific model parameters have clear across subject BOLD correlates, while other model parameters, as well as more standard behavioral measures do not. We also find that examining BOLD in terms of the expected decision-variable leads to a specific pattern of activation that would not be otherwise possible to extract.

It is somewhat surprising that when a stimulus is presented at a location irrelevant to the required response, that this location should feed into the decision process. This is what spatial conflict, and specifically the Simon effect^{1,2} demonstrates: the average response time (RT) of subjects is increased, and accuracy decreased if stimuli are presented in a lateralized manner so that they are incongruous with the response information that they represent symbolically.

The Simon effect deviates from the behavior that would be predicted by a simple application of the drift-diffusion model (DDM)^{3–5}. This deviation appears due to subjects making use of task-irrelevant information, and manifests in the shape of subjects' RT distributions^{6–9}. In recent years sequential sampling models (SSM) for conflict^{10,11} and the Simon task specifically¹² have been developed which fully capture RT distributions and accuracy.

The explained operation of the Simon effect SSM (SE-SSM¹²) within a conflict trial is rooted in models of cognitive control and prior models of the Simon effect^{13,14}. The SE-SSM, which we make use of here extends the SSM in two specific ways. It introduces an initial bias to the decision-variable, which more traditionally would be thought of as activation of an automatic¹⁵ pathway; and it also adds a conflict counteraction term, which acts to boost the decision-variable in conflict trials. From the perspective of model fitting, both of these parameters were shown to be essential in previous work¹². This is because while a bias term captures the increased mean response time in incongruent trials, the conflict counteraction term is needed to capture a compensatory decrease in the standard deviation of the RT - a key signature of the Simon effect. We identified the conflict counteraction term with a model driven approach, and so make no strong claims regarding its psychological significance, although we note that we have previously considered it to be involved in the deployment of attention.

¹Department of Biomedical Engineering, Columbia University, New York, NY, 10027, USA. ²Data Science Institute, Columbia University, New York, NY, 10027, USA. *email: j.mcintosh@gmail.com

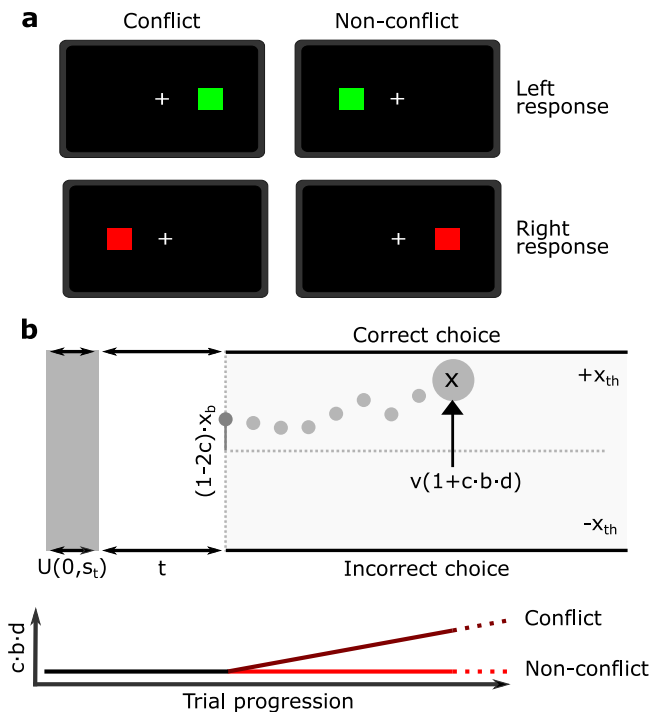


Figure 1. Task structure and model. **(a)** Subjects are presented with a green or red square. A green square indicates that the subject should make a response with their left hand while a red square indicates that the subject should respond with their right hand. Conflict trials are those where the side to which the square is presented to is inconsistent with the required response. **(b)** Top: Decision-variable to bound in the SE-SSM for a non-conflict trial. Parameters are explained in Eqs. 1–3. Bottom: Influence of conflict-counteraction on drift during conflict trials ($c = 1$). Variable d increases linearly during the trial, and is scaled by the fitted parameter b . During non-conflict trials ($c = 0$) conflict-counteraction does not impact drift. Decision-variable to bound in the SE-SSM for a non-conflict trial.

Such models lend themselves to enhance our understanding of the Simon task¹⁶ and enable new analysis in neuroimaging studies¹⁷. Previous studies have used fMRI to investigate the Simon effect^{18–23}, and while some of these have used analyses derived from Simon effect model considerations, none of these models could account for full RT and accuracy distributions. These previous analyses have however highlighted potential regions of interest such as dorsolateral prefrontal cortex (DLPFC), pre-supplementary motor area (pre-SMA), and anterior cingulate cortex (ACC)¹⁷. The power of the SSM is not only that it can model behavior, but that it is cast at a level of abstraction where its specific components are directly interpretable under the assumption that it captures some core principles of decision making used in the brain.

In this work we develop a model fitting framework which can be used to fit non-standard SSMs. We use this framework to fit the SE-SSM and show that it captures response time and accuracy of subjects that performed a Simon task. We go onto make use of the model fits by directly relating its parameters to fMRI data, and finally we exploit the high-temporal resolution of the model to decompose the BOLD signal.

Methods

Data source and task. Data analyzed in this manuscript was obtained from the OpenfMRI database under an ODC Public Domain Dedication and License (PDDL). Its accession number is ds000101²⁴. We reproduce here relevant details of the data, and refer the reader to the OpenfMRI database for a full description. Healthy adults ($n = 21$, male: 12, female: 9, age mean: 30, age std: 7) were scanned in a Siemens Allegra 3.0 T scanner, with a standard Siemens head coil, located at the NYU Center for Brain Imaging. 151 contiguous echo planar imaging (EPI) whole-brain functional volumes were obtained (TR = 2000 ms; TE = 30 ms; flip angle = 80, 40 slices, matrix = 64×64 ; FOV = 192 mm; acquisition voxel size = $3 \times 3 \times 4\text{mm}^3$) during each of the two sessions. A high-resolution T1-weighted anatomical image was also acquired using a magnetization prepared gradient echo sequence (MPRAGE, TR = 2500 ms; TE = 3.93 ms; TI = 900 ms; flip angle = 8; 176 slices, FOV = 256 mm).

As in a typical Simon task, subjects were presented a box to either the left or right hemifield and were asked to respond to its color. Upon presentation of a red box, subjects were asked to respond with their right index finger, while upon presentation of a green box they were asked to respond with their left index finger (see Fig. 1a). Two sessions were conducted per subject, with each session composed of 48 congruent, 48 incongruent trials and 24 null trials of 2.5 s. In the null trials, a stimulus was not presented and no response was necessary.

Simon effect model and fitting. *Model.* In previous work (see supplementary material), a SSM was proposed based on the the exact stimulus presentation (composed of visual hemifield, and color) and response (left

or right hand). As this information is not present in the available dataset, here we use a mathematically equivalent, accuracy coded form (Fig. 1b) where the model is defined in terms of a single stimulus conflict parameter, and consider the response only in terms of whether it was correct or incorrect (Eqs. 1–3):

$$\Delta x = v \cdot (1 + c \cdot b \cdot d)\Delta t + s\xi_1\sqrt{\Delta t} \quad (1)$$

$$x_0 = x_b(1 - 2c) \quad (2)$$

$$t^{eff} = t + \xi_2 s_t \quad (3)$$

The variable x denotes the decision-variable that builds to a threshold x_{th} . The stimulus conflict parameter c is set to take a value of 0 when there is no conflict present (i.e. when the stimulus is presented to the visual hemifield matching the required response hand), or 1 otherwise. Parameters v , s , t and x_0 correspond to the drift, noise, non-decision time and starting point bias, while ξ_1 represents the Gaussian diffusion process. The effective non-decision time on a given trial t^{eff} is defined as the sum of a base non-decision time t and a uniform noise process ξ_2 scaled by the noise in the non-decision time term s_t . Parameters that are specific to the Simon task are b , a conflict counteraction component that enhances the effective drift on conflict trials as the trial duration d increases (consequently reducing RT standard deviation). Note that the parameter d is a measure of the instantaneous time (measured in seconds) as a trial progresses, and this parameter is not fitted (see Fig. 1b). However, its influence is scaled by the parameter b . The parameter x_b captures the initial response bias caused by the Simon effect. Ultimately, parameters fitted for each subject are v , x_{th} , t , s_t , b and x_b .

Model fitting. Individual subject models were fit by using our SSM fitting framework (see supplementary material) which minimizes the negative log-likelihood by using Matlab's generalized pattern search algorithm²⁵. In order to generate the probability density for a specific set of parameters and conditions we resorted to encoding the probabilistic model decision-variable updates over time into a transition matrix (see supplementary methods and supplementary Fig. S1). We relied on such a method so that we could make use of likelihoods for model comparison, as well as it being relatively faster to evaluate (and consequently fit) than methods requiring the full simulation of individual decision-variable traces.

The transition matrix is fixed on non-conflict trials, and varies in a time dependent manner in conflict trials because of the $b \cdot d$ term. For each probability density evaluation, we discretized an initial decision-variable vector x so that it would span $\pm(x_{th} - 5s\sqrt{\Delta t})$, a range large enough to practically avoid being exceeded in a single step of the decision-variable. The x entry equivalent to the closest x_b value is set to one. The transition matrix is then iteratively defined for the conflict trials, or initially defined for the non-conflict trials by computing the expected Gaussian distribution for the decision-variable on the following trial for all possible current values of the decision-variable. Entries in the transition matrix outside of the decision thresholds are set to zero, except for where the transition leads to itself (i.e. diagonal entries), in which case they are set to one. The cumulative density function is then calculated by summing the iterated output above and below the model threshold. In order to generate the final density function, the cumulative density is numerically differentiated. We verified that these model fitting procedures worked sufficiently well to recover simulated parameters (see supplementary Figs. S3 and S4). Model fits were randomly initialized from the following set of distributions: $v \sim G(3, 1)$, $x_{th} \sim G(0.6, 0.15)$, $t \sim G(0.3, 0.075)$, $b \sim HN(2.5)$, $x_b \sim G(0.1, 0.06)$, $s_t \sim U(0, 0.15)$. Where, $G(x, y)$ represents the Gamma distribution parameterized by mean and standard deviation, $HN(x)$ represents the half Normal distribution parameterized by the standard deviation, and $U(a, b)$ represents the Uniform distribution parameterized by its two limits. The parameter s was fixed to one for all cases.

We initially fitted all our models with the decision-variable resolution set to 0.01, and repeated this five times to confirm that our procedure was consistently discovering global minima. After this procedure, we performed one further pass using the previous best model as a starting point with the decision-variable resolution set to 0.005 for more accurate parameter estimates.

BOLD processing. *Pre-processing.* Standard processing was implemented in FMRIB's Software Library (FSL)²⁶. Separation of non-brain tissue was performed using Brain Extraction Tool (BET)²⁷ with robust brain center estimation. Functional images were registered linearly to subject's structural image (BBR algorithm) and non-linearly (FNIRT, 12 DOF, 10 mm warp resolution) to a standard Montreal Neurological Institute (MNI) brain template. Functional data was high-pass filtered (0.01 Hz), and spatially smoothed with a 5 mm full-width half-maximum Gaussian kernel. The FEAT tool within FSL was used for event-related fMRI analysis. Stimuli were modeled as 1 s long step functions to match stimulus presentation length, and motion parameters (generated by MCFLIRT)²⁸ were added to the model as confounds of no interest. FEAT was set to use a canonical double-gamma haemodynamic response function to convolve with our regressors of interest.

Model driven across subject BOLD estimates. We hypothesized that if the SE-SSM is describing core components of the Simon effect that we may be able to see correlations between average stimulus activation across subjects and Simon task specific parameters (b and x_b). To investigate this, our analysis begins by application of a fixed effects general linear model (GLM) with a single set of fixed effect regressors generated by all stimuli for correct responses. This analysis is carried out at first (trial) level and extended to second (session) level, yielding GLM weights, which are then transformed into MNI space. MatlabTFCE²⁹ was used to generate p -values corresponding to each voxel after a mask excluding brain-stem and white matter was applied. Specifically for our application, MatlabTFCE calculates the correlation coefficients between GLM weights and parameter values of interest

across subjects for each voxel and then applies the Fisher z -transformation. A threshold-free cluster enhancement (TFCE)³⁰ transformation (default parameters were used: $E = 0.5$, $H = 2$) is then applied to these values, and the absolute value is taken (the absolute value is required for the two-tailed test). Generation of voxel p -values is then carried out by permutation. The permutation works as follows: at every iteration the z -transformed correlation is calculated from the parameters with respect to shuffled GLM weights, and a maximum z -value across voxels is calculated. For each voxel a count normalized by the total number of permutations performed is incremented every time the maximum z -value exceeded the real z -value, resulting in a maximal statistic permutation family-wise error rate (MSP FWER) corrected two-tailed p -value for every voxel. The p -value map is then split into two dependent on whether the real z -value is positive or negative, in order to generate a positive-tailed p -value map and a negative-tailed p -value map, with the complimentary voxels set to a p -value of 1.

To demonstrate that the model based analysis provides insights not accessible by simply examining the RT, we repeated the same analysis using the average RT across subjects, as well as the average difference in RT between congruent and incongruent trials across subjects.

Within trial BOLD estimates. We proceeded to investigate whether patterns of BOLD activation correspond to the development of the expected decision-variable (DV), a method similar in principle to the EEG-fMRI analysis used by Philiastides and Sajda³¹. In order to generate the expected decision-variable for each trial, we begin by simulating the fitted SE-SSM. Unlike the methods explained in the model fitting procedure, in this case the decision-variable is explicitly simulated at every time step until a threshold is reached. This process is repeated 7.5×10^5 times with the full decision-variable trace being saved, together with corresponding RT, and choice (correct or incorrect). For each real trial, we then match all simulated trials with the same choice and RT (up to a resolution of 25 ms), align them to their threshold crossing and average in order to generate the expected decision-variable trace over time. Once the decision-variable trace has been defined for all trials and for all subjects, we extract the average decision-variable in time windows stepping back from the response time. We set the window width to be 50 ms, starting centered at -25 ms and moved it back in 50 ms steps. The expected decision-variable was then extracted from a specific time window, z -scored and fed into a first level GLM which we term 'DV-GLM'. These DV based regressors were designed with 100 ms duration, centered on their extraction time³². This GLM additionally takes as its first feature, events corresponding to correct trials, and we carried it through to second level and third level (across subjects) with a mixed effects GLM. Clusters were thresholded at $z > 2.5$ and corrected for multiple comparisons using Gaussian random field theory (GFT) at $p < 0.05$. We repeated this process for each time window, generating an activation map for each.

Traditional BOLD analysis. We also conducted a more traditional BOLD analysis where we carry over first and second level GLM analyses to the third level (across subjects) with a mixed effects GLM. Clusters were thresholded at $z > 2.5$ and corrected for multiple comparisons using GFT at $p < 0.05$.

GLMs for this analysis were run with various combinations of variables of interest. A 'simple GLM' was initially implemented to test for whether there is a detectable response specific to incongruent stimuli, by using all stimuli as events for a first regressor and only incongruent stimuli as a second regressor. In this analysis, the coefficient matching the second regressor is then indicative of the difference in activation between congruent and incongruent stimuli. In the 'RT GLM', we proceeded to investigate whether the stimulus BOLD response is modulated by the RT, by using all stimuli as the first regressor, and the log-transformed and then z -scored RT for the second regressor.

Results

Simon effect is present in behavior. The Simon effect is visible in Fig. 2a, and we confirmed its presence by performing a Wilcoxon signed-rank test on subject RT averages across conditions ($p = 0.0006$, $n = 21$, $z = 3.42$). The average (mean) difference in RT between conflict and non-conflict trials was $25.7 \text{ ms} \pm 5.7 \text{ ms}$ (SEM). Standard deviations of the RT were also different between conflict and non-conflict ($p = 0.0087$, $n = 21$, $z = 2.62$), a trait common to the Simon effect, and potentially indicative of the involvement of cognitive control. As expected, accuracies in conflict trials were decreased with respect to non-conflict trials although the overall accuracy in both conditions was very high (non-conflict: $97.3\% \pm 0.5\%$ (SEM), conflict: $95.6\% \pm 0.5\%$ (SEM), Wilcoxon signed-rank $p = 0.0172$, $n = 21$, $z = 2.38$). The decreased accuracy in conflict trials with respect to non-conflict trials is driven by early responses (Fig. 2b,c), an observation which has been incorporated into Simon effect models.

Model captures Simon effect. As an informal assessment of the model fit quality for the SE-SSM, we re-sampled from it for each subject, and analyzed the data in the same way that we analyzed the behavioral data (c.f. Fig. 2a–c to Fig. 3a–c). The model fits demonstrate the Simon effect (Wilcoxon signed-rank test on subject RT averages, $p = 0.0006$, $n = 21$, $z = 3.42$). Visual inspection indicates a good match, the main difference being the across-subject variability which has a negligible within-subject component for our model. The average (mean) difference in RT between conflict and non-conflict trials was $29.2 \text{ ms} \pm 6.0 \text{ ms}$ (SEM). Standard deviations of the RT were also different between conflict and non-conflict (Wilcoxon signed-rank, $p = 0.0046$, $n = 21$, $z = 2.83$), and accuracies in conflict trials were decreased with respect to non-conflict trials (non-conflict: $98.3\% \pm 0.4\%$ (SEM), conflict: $96.6\% \pm 0.4\%$ (SEM), Wilcoxon signed-rank $p = 0.002$, $n = 21$, $z = 3.09$).

More formally, we computed Akaike information criterion (AIC) and weighted AIC (wAIC)^{33,34} for the SE-SSM (M4), and less complex candidate models. The less complex candidate models considered are: the basic SSM (M1), the basic SSM with the conflict counteraction parameter b (M2), the basic SSM with the bias parameter x_b (M3). The wAIC of a given model can be interpreted as the probability of that model being the best model among the candidate set³⁴.

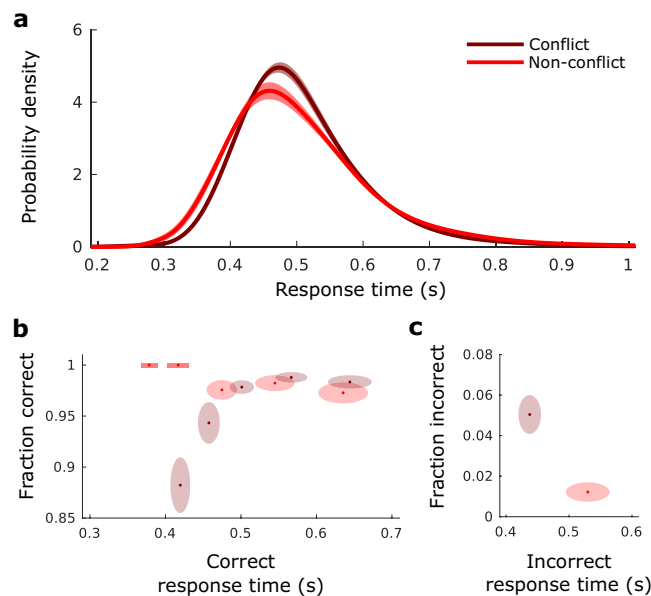


Figure 2. Behavioral data. (a) Subject averaged density estimate of correct RT for conflict and non-conflict responses. Simulated data is initially normalized at the individual subject level by dividing by subject standard deviation and then multiplying by grand mean RT, density estimation is performed by convolution with a Gaussian kernel. (b) Fraction of correct responses within response time percentile bins, plotted against central percentile RT. Percentiles bin edges were selected to be [0, 20, 40, 60, 80, 100]%. For example, the right most point shows the accuracy for data between the 80th and 100th percentile plotted against the 90th percentile response time. Ellipse width signifies SEM in RT across subjects, while ellipse height signifies standard-error in accuracy across subjects. (c) Same as (b) but for incorrect responses, a single percentile of 50% was used due to the low number of incorrect responses.

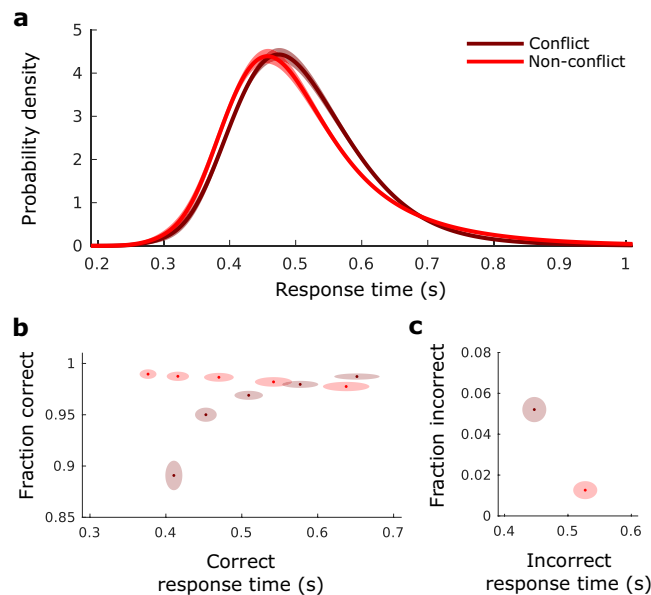


Figure 3. Model fitted simulated data for comparison to Fig. 2. Each subject model fit is used to produce 10^4 responses. (a–c) as in Fig. 2.

We found that candidate models M1 and M2 were not well-supported (largest wAIC for 2 subjects each). Models M3 and M4 on the other hand were both well-supported, with the largest wAIC coming from M3 for 8 subjects, and M4 for 9 subjects. This data is summarized in Fig. 4.

There are several points to note here: unsurprisingly, it seems that the addition of a bias term is essential to capture the Simon effect. On the other hand, the addition of the conflict counteraction term does not always clearly improve the model quality. While the variability between M3 and M4 may be partially due to sampling variability, it seems likely that some subjects simply behave more consistently with M3, while others behave more

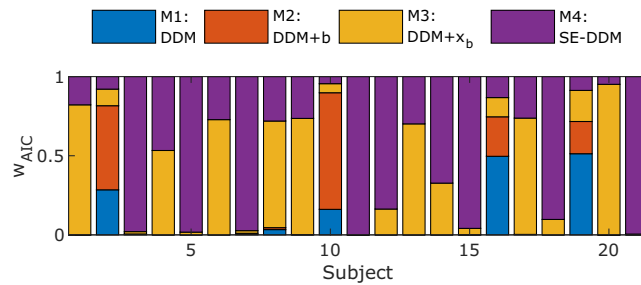


Figure 4. w_{AIC} for the basic SSM (M1), the basic SSM with the conflict counteraction parameter b (M2), the basic SSM with the bias parameter x_b (M3) and the SE-SSM (conflict counteraction parameter b and bias parameter x_b , M4). M3 and M4 are the best candidate models.

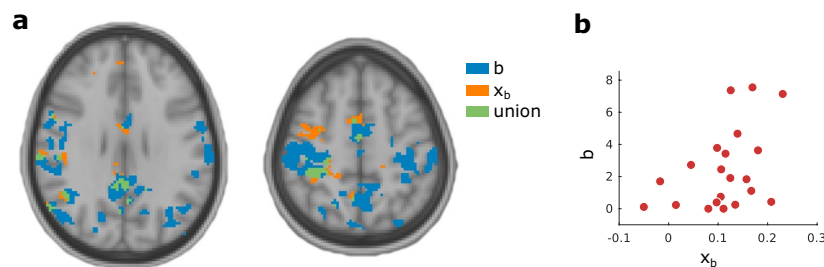


Figure 5. Regions where activity is related to SE-SSM parameters across subjects. (a) Regions where activity is positively correlated with ($p < 0.05$, MSP FWER corrected) the conflict counteraction parameter b are shaded in blue, while regions where activity is positively correlated with ($p < 0.05$, MSP FWER corrected) the bias parameter x_b are shaded in orange. Regions where both parameters highlight significant regions are shaded in green ($p < 0.05$, MSP FWER corrected). No negative correlations were found in this analysis. (b) While regions in a are mainly distinct, some overlap may be accounted for by the correlation between the bias and conflict counteraction parameter across subjects (Spearman's $CC = 0.48$, $p = 0.029$).

consistently with M4. Specifically, it should not be a surprise that there is variability in how subjects deploy conflict counteraction during the Simon task, with those that deploy it sparingly being captured best by M3.

Since the SE-SSM (M4) is the best fitting candidate in the largest number of our subjects, and it encompasses all other models (specifically, M3 is a special case of M4 when b is close to zero) we proceeded to analyze BOLD activity in terms of this model.

SE-SSM Across subject analysis. In this analysis we fitted a GLM with a single series of events corresponding to the stimuli. We then computed the correlation between our two task specific SE-SSM parameters (b , x_b) and the GLM weight maps across subjects. Activation, as assessed by permutation tests of the TFCE transformed correlations was present in several regions.

We repeated this correlation analysis with average subject RT, as well as the difference in subject RT between congruent and incongruent conditions (ΔRT) and found that no voxels were activated ($p < 0.05$, MSP FWER corrected). This demonstrates a clear benefit of the model based approach: RT and ΔRT are dominated by several sources of variance which are irrelevant to conflict or act to cancel each other out so that direct analysis may be inconclusive, while analysis of the underlying estimates of the contributions shows clear activation.

Correlations with bias x_b across subjects (Fig. 5a, orange $p < 0.05$, MSP FWER corrected) were present in multiple regions, most prominently and of particular interest the supplementary motor area (SMA), insular cortex, lingual gyrus, precuneus, posterior and anterior cingulate, the right precentral gyrus as well as the superior frontal and paracingulate gyrus and cerebellum.

Correlations with conflict counteraction b across subjects (Fig. 5, blue, $p < 0.05$, MSP FWER corrected) were most prominently present in SMA, insular cortex, lingual gyrus, precuneus, posterior and anterior cingulate, and the left and right precentral gyrus and cerebellum.

The partial overlap in brain activation (Fig. 5, green) related to the SE-SSM parameters led us to investigate whether across subjects, the parameters b and x_b are correlated. Indeed, as shown in Fig. 5b there does appear to be some correlation between the two parameters.

SE-SSM within trial analysis. Given the RT and whether the choice was correct on a given trial, we estimated the progression of the expected decision-variable (Fig. 6). We then used these values averaged over successive 50 ms windows as part of a GLM (see work by Philiastides and Sajda³¹ for a variation of this method applied in EEG-fMRI) in order to extract details regarding the progression of information during a trial. We found clusters of BOLD activity ($p < 0.05$, GFP FWER corrected, not corrected for multiple comparisons across time) with the expected decision-variable extracted from windows centered at -25 ms and -175 ms. Early activation

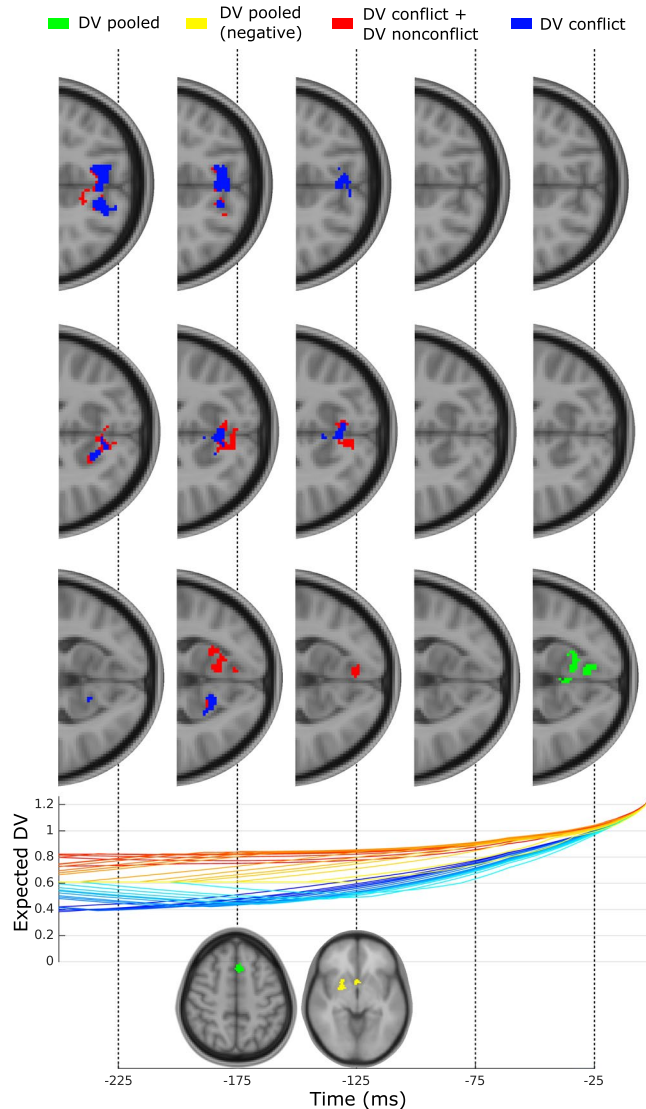


Figure 6. Expected decision-variable BOLD activation relative to response. Top: from left to right, transverse slices showing positive activation ($p < 0.05$, GFP FWER corrected, not corrected for multiple comparisons across time) from GLM with decision-variable taken from corresponding time. Activation shaded in green stems from GLM using decision-variable pooled from both conditions (we note additional activation at -175 ms, and a negative activation in yellow at -125 ms). Activation shaded in red and blue stem from GLM with the decision-variable split by condition, with red corresponding to a sum contrast, and blue corresponding to conflict only (non-conflict only did not yield any statistically significant regions). Bottom subplot shows an example decision-variable progression for one subject, and an axial slice at -175 ms. Orange: non-conflict, blue: conflict, shading dependent on RT rank.

at -175 ms includes the influence of the decision-variable jump (due to the bias term), smoothed by the 50 ms window, and may represent activation of the automatic pathway¹⁵ - itself a signature of conflict. It is therefore consistent, that the activated region at this stage is the paracingulate and superior frontal gyrus, neighboring both SMA and ACC. At -25 ms activation appears to be dominant in intracalcarine cortex and lingual gyrus.

To examine this further, we repeated our DV-GLM but split the expected decision-variable by whether it came from a congruent or incongruent trial. This analysis ensures that only the decision-variable variability within the congruent and, separately within the incongruent conditions can account for BOLD activation, thus removing the direct contribution of conflict. As well as assessing the statistical significance of the two decision-variable regressors, we also examined the contrast of their sum. While the non-conflict decision-variable did not yield any statistically significant BOLD activation, interestingly, both the conflict decision-variable and sum contrast did. The sum contrast activates regions largely overlapping to those activated by the single decision-variable GLM at -25 ms, although earlier with respect to the response and extending into cuneal cortex and precuneus.

Stimulus-response GLM analysis. We also performed a more traditional set of analyses which we termed 'Simple GLM' and 'RT GLM'.

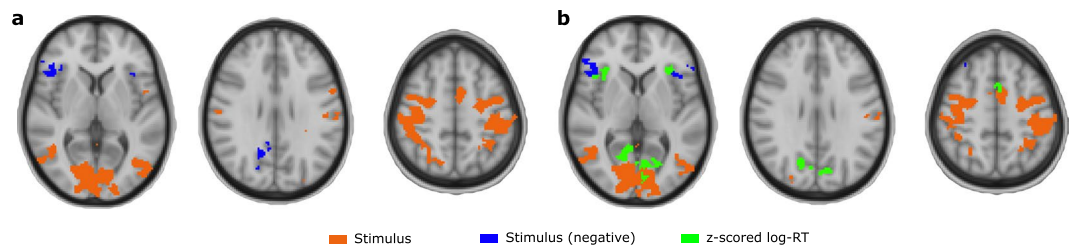


Figure 7. Standard GLM analyses. **(a)** Clusters for simple GLM. Orange clusters represent positive activation in response to stimulus, while blue clusters represent negative activation in response to stimulus. No additional clusters are present in response to an incongruent event. **(b)** Clusters for RT GLM. Orange and blue clusters represent activation in response to stimulus, similar to **(a)**, while green clusters represent activation dependent on z-scored log-RT events. No negative activation in relation to z-scored log-RT was found.

The simple GLM consists of modeling stimulus presentation as a single series of events, with incongruent events as an additional series. The stimulus presentation appears to positively activate clusters in the visual areas (large sections of occipital and lateral occipital cortex), response related regions such as the SMA, pre- and post central gyrus, the left central opercular cortex, as well as cerebellum. Negative activation was also visible in this analysis in insular cortex (bilaterally), middle temporal regions and precuneus.

In contrast, when examining coefficients coding for incongruence relative to stimulus presentation no clusters were present. This suggests (as has been found in previous work²¹) that the activation present due to conflict is subtle and may require more sophisticated analysis as we have presented here.

While incongruent stimuli may not directly lead to clear activation, we considered that RT may in part reflect the process of conflict detection. To test this, we performed a GLM with one event variable related to stimulus as before, and another event variable set to be the z-scored log-RT. The stimulus event clusters for this analysis are broadly similar to those presented for the previous analysis (stimuli and incongruent events, compare Fig. 7a to Fig. 7b). The z-scored log-RT event however appears to elicit activity in ACC, as well as bilaterally in insular cortex, and occipital-parietal areas ‘upstream’ to those visible for the stimulus event. These regions include intra/supracalcarine cortex and superior lateral occipital cortex.

Discussion

In this study, we consider the Simon effect to emerge out of three main components: (1) a representation of the choice relevant stimulus (color), stimulus laterality, and conflict with respect to choice; (2) evidence accumulators for each choice wherein their relative activation is the decision-variable; (3) cognitive control to appropriately distribute the computation across these components. We found that the SE-SSM which incorporates these components yields parameters which correlate with stimulus BOLD activation, while equivalent analysis based on RT does not.

Interpreting the SE-SSM in light of previously developed cognitive models^{6,13,14} we suggest that the decision-variable should be considered as equivalent to the relative activation of response accumulators, with the bias term representing the relative strength of the automatic pathway, and the conflict counteraction term linked to attention, monitoring and regulation. Across subject activation may depend on generic traits of the subject and should correspond to global parameter settings that project to regions where an individual trial is being processed, or targets of such regions. We also used the model to analyze the within BOLD correlates of the evolution of the expected decision-variable.

Because the RT distribution of the SE-SSM cannot be written analytically, in previous work we resorted to minimization of the χ^2 statistic³⁵ calculated from binned RT data which is generated by direct simulation of the SE-SSM process. Additionally, model comparison was done by examining the χ^2 statistic and application of cross validation, leading to a computationally expensive procedure. In this work, we developed a SSM framework to fit non-standard extensions of the SSM more efficiently. To do this, we encode the transition matrix of the SE-SSM and propagate this forwards in time in order to generate the model likelihoods. Such a method allows for a substantial speed up in model fitting so that we could be confident that we were avoiding local minima by repeating simulations. The method also directly allows us to use existing model comparison techniques without resorting to further simulations. We also note that the model estimation procedure employed, including the calculation of likelihood functions from a non-standard SSM could lend themselves to being incorporated into hierarchical SSM frameworks such as the HDDM³⁶.

We found that SE-SSM in general provided a good explanation for the behavioral data. However, the separation between the SE-SSM and SSM including conflict dependent bias was not as clear as we were expecting, with some subject models favoring the simpler form. Indeed, when applying the current methods to data from McIntosh *et al.*¹² (data not shown), we find that the SE-SSM is strongly preferred over all other candidate models (as had been previously shown using cross-validation). We hypothesize that this inconsistency may be due to the relatively weaker maximum trial time constraints placed on subjects in the current data set, providing less necessity for a strong counter active role of conflict counteraction on bias, however further experiments will be required to investigate this.

Using a GLM with all events, and additionally incongruent events as regressors, only yielded a statistically significant pattern of activation in response to all events. Performing a similar GLM with RT as opposed to

specifically modeling incongruent stimuli yielded a statistically significant ACC cluster. However, it is not clear whether this should be directly related to conflict as the RT ultimately is dependent on a variety of factors. This highlights a potential advantage of the model based approach: decomposition of accuracy and RT into model parameters acts to isolate concepts of conflict counteraction and bias which can then be used to interpret the data.

The ACC which we find to be clearly activated in our across subject analysis has been implicated in conflict monitoring³⁷ and resultant specification of a control signal¹³, which we hypothesise should be related to conflict counteraction. It is also generally found to be activated in several previous Simon effect fMRI studies^{18–20,38}, and is known to be activated with insular cortex³⁹. Aside from sensorimotor areas, other clear regions of activation are precuneus and the posterior cingulate. In previous work, we ascribed the concept of attention to our conflict counteraction model parameter b , and so it is interesting to note that precuneus has also been found to be active in previous Simon effect studies, and has been implicated in shifting attention to different spatial locations⁴⁰. However, we currently make no strong claims regarding specific theories of attention⁴¹.

While activation related to the bias term was present in frontal and paracingulate gyri, we found that its activation largely overlapped with the activation of the conflict counteraction parameter. This would be partially predicted by the correlation between conflict counteraction and bias, which we speculate may be due to a strategy by subjects with a strong bias, compensating by enhancing their conflict counteraction. This study, and previous work¹² suggests that two task specific parameters are required to fully capture behavior in the Simon task, so we hypothesize that the overlap of the brain activity is due to the interwoven and distributed nature of the decision making process. We cannot however rule out that there is a more parsimonious model structure that would allow the Simon effect to be captured with a single variable which itself optimally reflects brain activity. This question will need to be resolved in future studies where the conflict counteraction parameter is modulated experimentally.

Within a trial, the consistent activation of a single region shows a location where a decision-variable could be directly represented, or where components required for its calculation are represented. For example, regions activating during the early representation of the decision-variable progression could be related to automatic activation, while later activation could be related to evidence integration and conflict counteraction. An example of our interpretation for the specific process that is occurring during a Simon effect trial is as follows: a left indicating stimulus is presented to the right hemifield. An automatic pathway is activated, and feeds into a response accumulator. This pathway is quickly inhibited indirectly, or its activity is inherently transient so that while an impact has been made on the current state of the accumulators, further impact is reduced. We model this as a delta function at a moment corresponding to the non-decision time - although a more natural process starting somewhat earlier is likely to occur in reality. The non-automatic, task specific process is activated less quickly - and begins to feed into the pre-motor accumulators once the automatic pathway is no longer having a novel impact. On some occasions the inhibition is not fully deployed, or noise in the decision process acts to initiate an action at this early stage. Usually however, this is not the case and evidence from the non-automatic pathway gradually pushes the correct response accumulator towards its threshold. Due to time constraints of the task, or the inherent cost of time however, a compensation mechanism takes hold to reduce the influence of the initial bias. Such a mechanism should make use of the representation of the decision-variable in terms of the initial stimulus laterality: a measure of conflict which changes as the trial progresses. Monitoring of this conflict could in turn be used to accelerate the decision-variable towards a bound.

In our within trial analysis, we attempt to examine the progression of the decision with high temporal resolution despite fMRI's low temporal resolution using a method inspired by EEG-fMRI analysis³¹ and the concept of examining the expected decision-variable⁴². We found clear representation of the decision-variable early and late relative to the response. Early activation (175 ms prior to response) appears to be in paracingulate and superior frontal gyri which neighbors SMA and ACC. It seems possible that this corresponds to early detection of conflict, or deployment of conflict counteraction in response to conflict, which would be represented in the decision-variable at an early stage in a graded manner due to the smoothing effect of our temporal windows. We also detect late activation relative to response time (–25 ms) of intracalcarine cortex. This type of activation would be consistent with full representation of a decision-variable like feature prior to response, although its presence in early visual areas late in the decision is surprising. We were curious as to whether we might find a separate representation of the conflict specific decision-variable, and found that unlike the decision-variable taken from all trials, activation is more strongly represented in precuneus. One interpretation of this progression is consequently that posterior medial frontal cortex monitors conflict³⁷ for the purposes of modulating an autonomic pathway¹³, or up-regulating task relevant (color) information in visual areas⁴³. In general, our time resolved BOLD analysis method may hold promise for extracting temporal information from slowly changing BOLD responses within a trial, rather than relying on across trial changes induced by task conditions⁴³. In particular, we believe this may be the case with appropriately designed paradigms that allow for a complex progression of the expected DV allowing GLMs conducted at different time points to be uncorrelated.

A weakness of the current study, particularly relevant for the time resolved analysis stems from the data being cast in terms of conflict and accuracy, rather than stimulus and response: stimulus and response laterality influence the decision-variable in different ways than conflict and accuracy producing potentially different patterns of activation. For example, we would expect the early stimulus-response decision-variable to be lateralized in the occipital region, while the late decision-variable may be lateralized in the motor or pre-motor region. In this study, we detect early ACC activation in our within-trial analysis, potentially corresponding to conflict detection. While it may be the case that conflict detection is occurring early in the trial, an alternative explanation is that early expected decision-variable is just an indirect signature of conflict (much like RT is), which is itself only represented later in a trial. This possibility can be analyzed more carefully with exact stimulus and choice information, where conflict is not directly confounded with the early decision-variable. We also note that the GLM analyses conducted across time were not corrected for multiple comparisons relative to each other, so the possibility of type II error in the interpretation of some of these clusters is inflated. In future work, we intend to

address these points, and note that improved experimental conditions may also allow us to analyze the across trial intermediate temporal domain and influence of conflict history in the context of the proposed model (see supplementary material Figs. S4 and S5).

We introduced a novel method to fit our conflict based Simon effect model (SE-SSM), and have demonstrated that the SE-SSM captures behavioral data. The two conflict specific model parameters it introduces have clear across-subject BOLD correlates, and, furthermore we have shown that SSM type models can be used to decompose the BOLD activation via the lens of what would be expected at different time points during a trial.

Data availability

The datasets analyzed during the current study are available from the OpenfMRI database under an ODC Public Domain Dedication and License (PDDL). Its accession number is ds000101²⁴. Our model fitting framework is available at <https://github.com/jrmxn/malleable-ssm>.

Received: 29 October 2019; Accepted: 16 February 2020;

Published online: 03 March 2020

References

- Simon, J. R. & Rudell, A. P. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *J. Appl. Psychol.* **51**, 300–304 (1967).
- Simon, J. R. The Effects of an Irrelevant Directional cue on Human Information Processing. In Reeve, R. & Proctor, T. (eds.) *Advances in Psychology*, vol. 65 of *Advances in Psychology*, 31–86 (North-Holland, 1990).
- Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
- Ratcliff, R. & McKoon, G. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Comput.* **20**, 873–922 (2008).
- Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion Decision Model: Current Issues and History. *Trends Cogn. Sci.* **20**, 260–281 (2016).
- De-Jong, R., Liang, C.-C. & Lauber, E. Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 731–750, <https://doi.org/10.1037/0096-1523.20.4.731> (1994).
- Pratte, M. S., Rouder, J. N., Morey, R. D. & Feng, C. Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophys.* **72**, 2013–2025 (2010).
- Schwarz, W. & Miller, J. Response time models of delta plots with negative-going slopes. *Psychon. Bull. & Rev.* **19**, 555–574, <https://doi.org/10.3758/s13423-012-0254-6> (2012).
- Servant, M., Montagnini, A. & Burle, B. Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cogn. Psychol.* **72**, 162–195 (2014).
- Ulrich, R., Schröter, H., Leuthold, H. & Birngruber, T. Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cogn. Psychol.* **78**, 148–174 (2015).
- White, C. N., Servant, M. & Logan, G. D. Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychon. Bull. & Rev.* **25**, 286–301, <https://doi.org/10.3758/s13423-017-1271-2> (2017).
- McIntosh, J. R. & Mehring, C. Modifying response times in the Simon task with transcranial random noise stimulation. *Sci. Rep.*, **7**, <https://doi.org/10.1038/s41598-017-15604-1> (2017).
- Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240, <https://doi.org/10.1016/j.neuron.2013.07.007> (2013).
- Ridderinkhof, R. Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychol. Res.* **66**, 312–323, <https://doi.org/10.1007/s00426-002-0104-7> (2002).
- Hommel, B. Spontaneous decay of response-code activation. *Psychol. research* **56**, 261–268 (1994).
- Salzer, Y., de Hollander, G. & Forstmann, B. U. Sensory neural pathways revisited to unravel the temporal dynamics of the Simon effect: A model-based cognitive neuroscience approach. *Neurosci. & Biobehav. Rev.* **77**, 48–57 (2017).
- van Maanen, L., Turner, B. & Forstmann, B. U. From model-based perceptual decision-making to spatial interference control. *Curr. Opin. Behav. Sci.* **1**, 72–77, <https://doi.org/10.1016/j.cobeha.2014.10.010> (2015).
- Peterson, B. S. et al. An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cogn. Brain Res.* **13**, 427–440, [https://doi.org/10.1016/s0926-6410\(02\)00054-x](https://doi.org/10.1016/s0926-6410(02)00054-x) (2002).
- Liu, X., Banich, M. T., Jacobson, B. L. & Tanabe, J. L. Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage* **22**, 1097–1106, <https://doi.org/10.1016/j.neuroimage.2004.02.033> (2004).
- Kerns, J. G. Anterior cingulate and prefrontal cortex activity in an fMRI study of trial-to-trial adjustments on the Simon task. *NeuroImage* **33**, 399–405, <https://doi.org/10.1016/j.neuroimage.2006.06.012> (2006).
- Forstmann, B. U. et al. Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: A model-based approach. *J. Neurosci.* **28**, 9790–9796, <https://doi.org/10.1523/jneurosci.1465-08.2008> (2008).
- Forstmann, B. U., van den Wildenberg, W. P. M. & Ridderinkhof, K. R. Neural mechanisms, temporal dynamics, and individual differences in interference control. *J. Cogn. Neurosci.* **20**, 1854–1865, <https://doi.org/10.1162/jocn.2008.20.122> (2008).
- Strack, G., Kaufmann, C., Kehrer, S., Brandt, S. & Stürmer, B. Anticipatory regulation of action control in a Simon task: Behavioral, electrophysiological, and fMRI correlates. *Front. Psychol.* **4**, <https://doi.org/10.3389/fpsyg.2013.00047> (2013).
- Kelly, A. M. C. & Milham, M. P. NYU Simon Task, revision: 2.0.0 OpenfMRI Accession number: ds000101 (2011).
- Kolda, T., Lewis, R. & Torczon, V. Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods. *SIAM Rev.* **45**, 385–482 (2003).
- Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208–S219, <https://doi.org/10.1016/j.neuroimage.2004.07.051> (2004).
- Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155, <https://doi.org/10.1002/hbm.10062> (2002).
- Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841, <https://doi.org/10.1006/nimg.2002.1132> (2002).
- Thornton, M. MatlabTFCE, <https://github.com/markallenthornton/MatlabTFCE> (2015).
- Smith, S. M. & Nichols, T. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44**, 83–98, <https://doi.org/10.1016/j.neuroimage.2008.03.061> (2009).
- Philiastides, M. G. & Sajda, P. EEG-Informed fMRI Reveals Spatiotemporal Characteristics of Perceptual Decision Making. *The J. Neurosci.* **27**, 13082–13091 (2007).
- Walz, J. M. et al. Simultaneous eeg-fMRI reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. *J. Neurosci.* **33**, 19212–19222 (2013).

33. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Autom. Control.* **19**, 716–723, <https://doi.org/10.1109/tac.1974.1100705> (1974).
34. Wagenmakers, E.-J. & Farrell, S. AIC model selection using akaike weights. *Psychon. Bull. & Rev.* **11**, 192–196, <https://doi.org/10.3758/bf03206482> (2004).
35. Ratcliff, R. & Tuerlinckx, F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. & Rev.* **9**, 438–481 (2002).
36. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. *Front. Neuroinformatics* **7**, <https://doi.org/10.3389/fninf.2013.00014> (2013).
37. Ridderinkhof, K. R. The role of the medial frontal cortex in cognitive control. *Science* **306**, 443–447, <https://doi.org/10.1126/science.1100301> (2004).
38. Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M. & Ullsperger, M. Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *J. Neurosci.* **31**, 1780–1789, <https://doi.org/10.1523/jneurosci.4299-10.2011> (2011).
39. Medford, N. & Critchley, H. D. Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. *Brain Struct. Funct.* **214**, 535–549, <https://doi.org/10.1007/s00429-010-0265-x> (2010).
40. Wenderoth, N., Debaere, F., Sunaert, S. & Swinnen, S. P. The role of anterior cingulate cortex and precuneus in the coordination of motor behaviour. *Eur. J. Neurosci.* **22**, 235–246, <https://doi.org/10.1111/j.1460-9568.2005.04176.x> (2005).
41. Petersen, S. E. & Posner, M. I. The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).
42. Selen, L. P. J., Shadlen, M. N. & Wolpert, D. M. Deliberation in the motor system: reflex gains track evolving evidence leading to a decision. *The J. Neurosci.* **32**, 2276–2286 (2012).
43. Egner, T. & Hirsch, J. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.* **8**, 1784 (2005).

Acknowledgements

We thank Anna Jasper and Tao Tu for comments. This work was partially funded by the Army Research Laboratory (ARL) under cooperative agreement number W911NF-10-2-0022, we acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010. We acknowledge the OpenfMRI project and NSF Grant OCI-1131441 (R. Poldrack, PI) for making this data accessible. Declarations of interest: none.

Author contributions

J.R.M. conceived and performed analysis. J.R.M and P.S. wrote manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-60943-1>.

Correspondence and requests for materials should be addressed to J.R.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020