RESEARCH ARTICLE

The posterior cerebellum and temporoparietal junction support explicit learning of social belief sequences

Qianying Ma[1] · Min Pu[1] · Naem P. Haihambo[1] · Kris Baetens[1] · Elien Heleven[1] · Natacha Deroost[1] · Chris Baeken[2,3,4,5] · Frank Van Overwalle[1]

Abstract

This study tests the hypothesis that the posterior cerebellum is involved in social cognition by identifying and automatizing sequences of social actions. We applied a belief serial reaction time task (Belief SRT task), which requires mentalizing about two protagonists' beliefs about how many flowers they receive. The protagonists' beliefs could either be true or false depending on their orientation (true belief: oriented towards and directly observing the flowers; or false belief: oriented away and knowing only prior information about flowers). A Control SRT task was created by replacing protagonists and their beliefs with shapes and colors. Participants were *explicitly* told that there was a standard sequence related to the two protagonists' belief orientations (Belief SRT task) or the shapes' colors (Control SRT task). Both tasks included a Training phase where the standard sequence was repeated and a Test phase where this standard sequence was interrupted by random sequences. As hypothesized, compared with the Control SRT task, the Belief SRT task recruited the posterior cerebellar Crus II and the temporoparietal junction (TPJ) more. Faster response times were correlated with less Crus II activation and with more TPJ activation, suggesting that the Crus II supported automatizing the belief sequence while the TPJ supported inferring the protagonists' beliefs. Also as hypothesized, compared with an *implicit* version of the Belief SRT task (i.e., participants did not know about the existence of sequences; Ma, Pu, et al., 2021b), the cerebellar Crus I &II was engaged less during initial training and automatic application of the sequence, and the cortical TPJ was activated more in processing random sequences.

Keywords Serial reaction time task · False belief · Explicit sequence learning · Cerebellum · Social cognition

Introduction

Successful social interactions require people to understand the mental states of others (e.g., beliefs, intentions), which is termed "*Mentalizing*" or "*Theory of Mind*" (ToM; Premack & Woodruff, 1978; Van Overwalle, 2009). Although the role of mentalizing is widely acknowledged by scholars in the field of social cognition, less attention has been given to associated implicit or explicit learning of social routines and sequences, which help to anticipate human interaction and cooperation, and so determine how smoothly and easily they may proceed. Indeed, social interactions often require people to continuously monitor and update what others may see, know, and believe. Advanced knowledge or correct anticipation of others' mental states may render interactions more efficient, such as during conversations (Mastroianni et al., 2021).

Although traditionally regarded as uniquely involved in motor automatization and execution, accumulating evidence

✉ Qianying Ma
  qianying.ma@vub.ac.be

✉ Frank Van Overwalle
  Frank.VanOverwalle@vub.be

[1] Department of Psychology, Center for Neuroscience, Vrije Universiteit Brussel, Pleinlaan 2, B -, 1050 Brussels, Belgium

[2] Department of Head and Skin, Faculty of Medicine and Health Sciences, Ghent Experimental, Ghent, Belgium

[3] Psychiatry (GHEP) Laboratory, Ghent University, Ghent, Belgium

[4] Department of Psychiatry, University Hospital (UZBrussel), Brussels, Belgium

[5] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

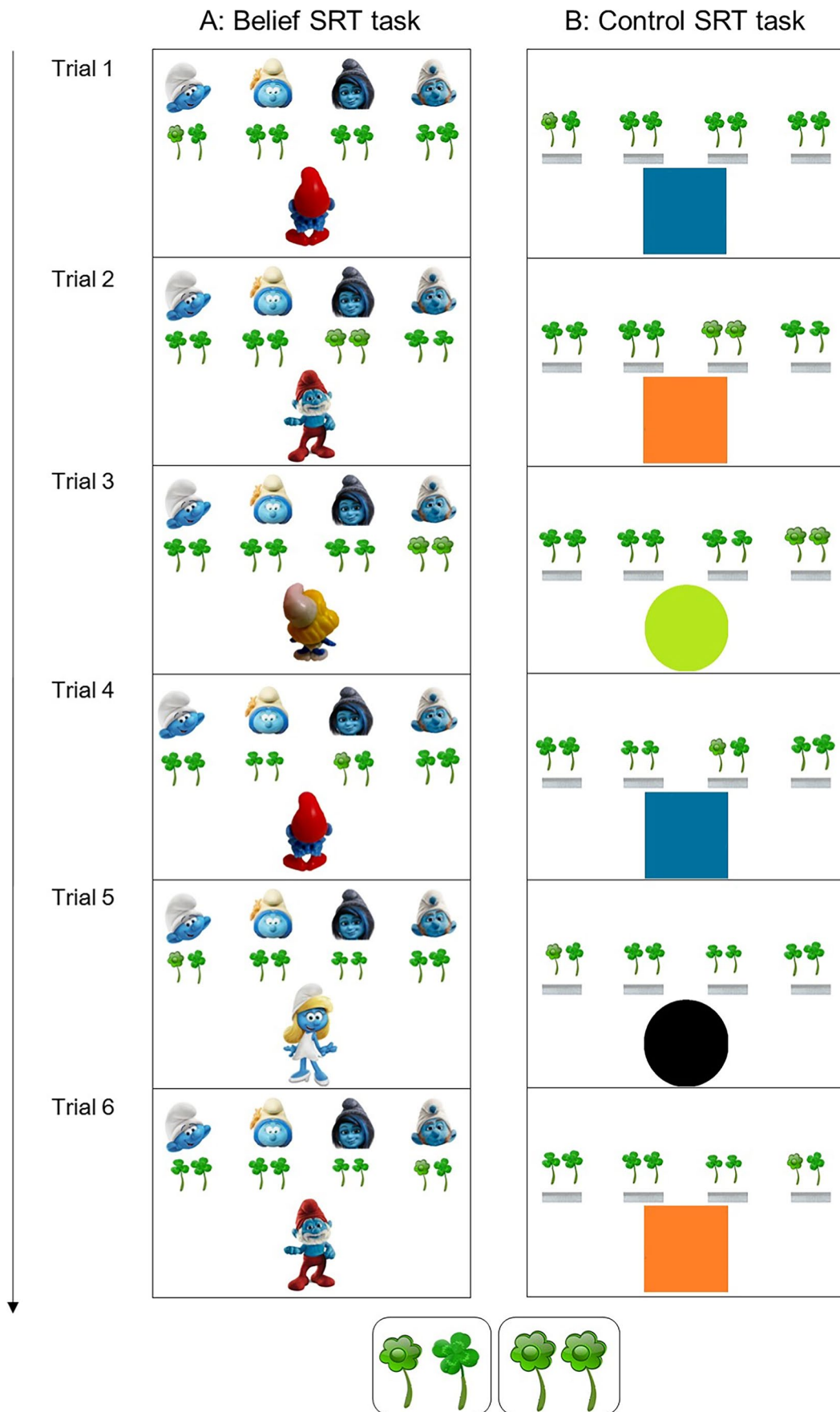## A: Belief SRT task

B: Control SRT task

◀**Fig. 1** Schematic example showing the first six trials of the Standard sequence in the explicit Belief SRT task (a) and the explicit Control SRT task (b). On each trial, participants had to report the number of flowers as seen by the protagonists (Papa Smurf or Smurfette in the Belief SRT task) or depending on the color variation of the shapes (square or circle in the Control SRT task). Belief orientations/color variations followed a Standard sequence. In the Belief SRT task, when the protagonist was oriented to the screen and could see the flowers (true trial), the number of target flowers had to be reported from the current trial; when the protagonist was oriented away from the screen and could not see the flowers (false trial), the number of target flowers had to be reported from the previous true trial from the same protagonist. Similarly, in the Control SRT task, a blue square or a green circle indicated that the number of flowers had to be taken from the current trial (= true trial), while an orange square or black circle (= false trial) indicated that the number of flowers had to be taken from the previous true trial with the same shape. The number of flowers was random (1 or 2), making the response unpredictable, and dissociating sequence learning from motor responses. Each trial was self-paced, with all stimuli remaining on screen for 3,000 ms until a response was given and was followed by a response-stimulus interval of 400 ms before the next trial started. [**Bottom Inset**] The inset shows an enlargement of the target stimulus, consisting of a pair of one or two flowers surrounded by clovers (as a distraction) of approximately the same shape and color. [**Trial 1 - 2**] To illustrate the instructions for the Belief SRT task, in Trial 1, there is one flower that Papa Smurf can see because he is oriented toward the screen, meaning that the correct response is 1. In Trial 2, there are two flowers. Because Papa Smurf is oriented away from the screen, he cannot see the number of flowers on this trial, hence he still thinks to have received one flower which he last saw on the previous (1st) trial. The correct response is thus again 1. In the Control SRT task, Trial 1, there is one flower. Because the color of the square is blue, the correct response is the observed number of flowers, or 1. In Trial 2, the square is orange, so participants must report the number of flowers from the blue square on the previous (1st) trial. The correct response is thus, again, 1

in the past decade suggests that the cerebellum also is engaged in detecting and learning repetitive patterns of pure mental states, including social cognition (Ito, 2008; Leggio & Molinari, 2015). A meta-analysis by Van Overwalle et al. (2014) on more than 350 fMRI studies related to social cognition revealed robust activation in the posterior cerebellum during mentalizing tasks. Importantly, these cerebellar activations largely overlapped with the default/mentalizing network identified by Buckner et al. (2011; Van Overwalle et al., 2015). Another recent meta-analysis found that a great majority (74%) of approximately 200 studies with activation in the posterior cerebellar Crus II revealed task processes involving mentalizing or self-related emotion attribution (Van Overwalle, Ma, & Heleven, 2020a). Recent fMRI studies further indicated that the posterior cerebellum was highly activated when social tasks required the generation of social action sequences, such as giving the correct chorological order of stories involving beliefs (Heleven et al., 2019), memorizing the order of actions implying traits of other people (Pu et al., 2020; Pu et al., 2021), or predicting sequential actions of persons based on their traits (Haihambo et al., 2021). Together, these studies suggest that the posterior

cerebellum, especially Crus I and II, plays a critical role in identifying the sequence of social actions while inferring the mental state of other persons and may aid in anticipating and inferring others' mental states during dynamic interaction.

While in the above-mentioned studies, social action sequences were generated or learned explicitly (i.e., with full awareness that sequences were present), a recent behavioral study investigated sequence learning involving others' beliefs implicitly (i.e., with little or no awareness; Ma, Heleven, et al., 2021a). In this study, a novel belief serial reaction time task was used (*Belief SRT task*). As depicted in Fig. 1a, participants saw one of two protagonists (Papa Smurf or Smurfette) at the bottom of the screen. This protagonist who was either orientated towards or away from the screen so that he or she could or could not see the flowers offered to them. To report how many flowers Papa Smurf or Smurfette thought were offered, participants had to mentalize regarding what each protagonist knows and believes. Specifically, participants were told that, if protagonists were oriented towards the flowers, they held a true belief about reality. Conversely, if they were oriented away from the flowers, they were not aware of any changes and therefore held a false belief. A nonsocial Control SRT task also was created (Fig. 1b), where shapes replaced protagonists and colors replaced beliefs, but otherwise, the structure of the task remained identical.

In both SRT tasks, there was a standard sequence involving a fixed order of the protagonists' belief orientations (Belief SRT task) or the shapes' colors (Control SRT task). In particular, there was a Training phase where this standard sequence was repeated, followed by a Test phase where this standard sequence was interrupted by random sequences. Although participants were not informed about the existence of a sequence (implicit instruction), the study demonstrated that people can implicitly learn the standard sequence in both Belief and Control SRT tasks, as revealed by faster reaction times to the standard sequence and slower reaction time to random sequences (Ma, Heleven, et al., 2021a).

A follow-up fMRI study using identical tasks and implicit instructions showed that, compared with the Control SRT task, the cerebellar Crus I was recruited during learning a sequence of beliefs (during Training) and detecting disruptions of this sequence (during Test), whereas the Crus II was activated when this belief sequence continued, but with occasional sequence disruptions (during Test; Ma, Pu, et al., 2021b). The temporoparietal junction (TPJ) also was activated when participants were inferring the protagonists' beliefs during initial learning of the standard sequence and detecting disruptions of it. This is consistent with the central role of TPJ in belief attribution (see meta-analyses by Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle, 2009). These experiments indicate that participants could implicitly learn a sequence based on other's

beliefs in a social context and that the cerebellar Curs I & II, and the cortical TPJ contributed to this implicit learning process.

Given the role of the cerebellum in automatization, it is generally assumed that the cerebellum is more engaged during implicit sequence learning compared with explicit learning, because explicit instructions possibly engage other neocortical areas, such as the prefrontal cortex, which may support initial learning. To illustrate, in a study by Morgan et al. (2020), patients with cerebellar degeneration had to respond to a target's spatial location. The authors found that compared with healthy participants, patients' "implicit sequence learning [was] impaired. However, for cognitive sequencing that could be accomplished using explicit strategies, the cerebellar [patients] performed normally." (pp. 222, Morgan et al., 2020). In another study, Taylor et al. (2010) used implicit and explicit instructions in a task that required participants to make ballistic-style reach movements to reach a target. Compared with healthy participants, cerebellar patients showed poor implicit learning (i.e., when they were not informed about a disturbance in their movement), but their learning improved when they were given explicit instructions on how to compensate for these disturbances. However, a recent study that included social contexts revealed impaired performance even when cerebellar patients were explicitly asked to give correct orders of stories requiring the understanding of other's beliefs, while they performed at normal levels for routine social or physical events (Van Overwalle, De Coninck, et al., 2019a). These findings raise two pivotal questions: (1) Given that the posterior cerebellum is preferentially involved in implicit learning of social sequences as demonstrated by Ma, Pu, et al. (2021b), does it also contribute to explicit social learning? (2) Is there a difference between explicit and implicit sequence learning in a social context?

To address the first question regarding whether the cerebellum is involved in explicit sequence learning in a social context, the present study used the Belief and Control SRT tasks from the previous study by Ma, Heleven, et al. (2021a). However, unlike that study in which participants were not informed about the presence of a standard sequence (implicit instruction; hereafter termed "implicit" SRT tasks), participants were now explicitly told that there was a standard sequence (explicit instruction; hereafter termed "explicit" SRT tasks). Similar to the previous study (Ma, Pu, et al., 2021b), we hypothesized that the posterior cerebellar Crus I & II and cortical TPJ will be more activated in the explicit Belief SRT task than in the explicit Control SRT task as these two areas are preferentially engaged in social contexts (**Hypothesis 1**).

To address the second question whether explicit and implicit social sequence learning differ, we compared the current explicit Belief SRT task to the previous implicit Belief SRT task (Ma, Pu, et al., 2021b). Because clinical studies have suggested stronger cerebellar involvement during implicit, rather than explicit, sequence learning (Morgan et al., 2020; Taylor et al., 2010), we hypothesized that, compared with the implicit Belief SRT task, the present explicit Belief SRT task may recruit less activation in the posterior cerebellar Crus, because the cerebellum is more engaged in implicit learning. Because participants in both implicit and explicit Belief SRT tasks received the same mentalizing instruction (i.e., how to use the protagonists' orientations) and thus engaged in mentalizing explicitly, we hypothesized a similar level of activation in the TPJ (**Hypothesis 2**).

Method

Participants

A total of 46 healthy, right-handed, Dutch-speaking participants were recruited. All of them had normal or corrected-to-normal vision and color perception. To avoid any carryover effects of sequence learning (Geiger et al., 2018), participants were randomly assigned between the explicit Belief and explicit Control SRT tasks. One participant was excluded because of excessive head movement (outlier scans >5 %), and one participant was excluded because of dizziness experienced during the experiment. Hence, data analysis was based on 22 participants in the explicit Belief SRT task (17 females, age 19-34 years, mean age 23.3 ± 4.1) and 22 participants in the explicit Control SRT task (18 females, age 18-32 years, mean age 24.1 ± 3.8). All participants gave written, informed consent with the approval of the Medical Ethics Committee at the University Hospital of Ghent. Participants were paid 20 euros, and transportation costs were reimbursed in exchange for their participation.

Stimuli material

The stimulus material and procedures in the current tasks are identical to those used in the previous implicit learning tasks (Ma, Pu, et al., 2021b) with the exception that participants were explicitly informed of the existence of a sequence in the current task.

Explicit Belief SRT task

In the explicit Belief SRT task (Fig. 1a), the target was one or two flowers, appearing in one of four horizontal locations, marked by four little Smurfs, on the top of the screen. The target flower(s) was presented along with clovers as distractors, which occupied other remaining locations, resulting in two plants on each location. The two protagonists, Papa Smurf and Smurfette, were each shown individually at the

bottom of the screen with their face orientated either toward or away from the screen.

Participants were instructed: "One of the four little Smurfs will give the flowers while Papa Smurf or Smurfette is watching (facing the screen) or not watching (facing you). Papa Smurf and Smurfette count the flowers they receive. Throughout the task you have to track how many flowers (1 or 2) that Papa Smurf or Smurfette thinks he or she will get. If they are turned with their back to the four Smurfs, you have to indicate how many flowers they (remember that they) received last time." To encourage participants to focus on belief sequences, they were explicitly informed about the kind of sequence they should search for: "WATCH OUT! In this task there is a fixed sequence of Papa Smurf and Smurfette and their orientations (towards or away from the screen). Try to find this sequence as that will make the task easier for you." (Best translation from Dutch). This explicit information avoided possible misunderstanding about other potential sequences, such as sequences based on flowers' location. However, participants were not informed about the exact sequence itself, so that the task involved learning rather than memorizing (Deroost & Coomans, 2018).

Explicit Control SRT task

The following changes were made in the non-social explicit Control SRT task (Fig. 1b). The target flower(s) was marked by four sidewalk boards instead of little Smurfs. The orientations of Papa Smurf (true and false) and Smurfette (true and false) were replaced by squares (blue and orange) and circles (green and black) respectively. Thus, the four distinct pictures used in the explicit Belief SRT task were replaced by four distinct pictures of colored shapes in the explicit Control SRT task.

Participants were instructed: "At the bottom of the screen is a square or circle. Throughout the task you have to follow how many flowers (1 or 2) there are at the blue square or the green circle. If the square is orange, you must report the previous number of flowers from the blue square. If the circle is black, you have to report the previous number of flowers from the green circle. Again, participants were explicitly informed about the kind of sequence they should search for, without informing them about the exact sequence itself: "WATCH OUT! In this task there is a fixed sequence of two shapes and their colors. Try to find this sequence as that will make the task easier for you." (Best translation from Dutch).

In both explicit Belief and explicit Control tasks, the number of flowers was randomly determined at every trial, leading to random motor responses, and a dissociation between the sequence of belief orientations/color variations and motor responses, while the sequence embedded in the two tasks and the instructions were structurally the same (Fig. 2b). Together, this was in all respects identical to the prior implicit Belief SRT task (Ma, Pu, et al., 2021b),

A: The experimental design of the Belief and Control SRT Tasks

Training phase					Test Phase																									
1	2	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
					Order 1																								
S	S	S	S	S	S	TR	RO	S	RO	TR	S	TR	RO	S	TR	RO	S	TR	RO	S	RO	TR	S	TR	RO	S	RO	TR	S
					Order 2																								
					S	RO	TR	S	TR	RO	S	RO	TR	S	RO	TR	S	RO	TR	S	TR	RO	S	RO	TR	S	TR	RO	S

B: Standard sequence in the Belief SRT task (repeated every 16 trials)

Protagonist	M	M	Fe	M	Fe	M	Fe	Fe	Fe	Fe	Fe	M	Fe	M	Fe	M
Belief Orientation	T	Fa	T	T	Fa	Fa	T	Fa	Fa	T	Fa	T	T	Fa	Fa	T
Prior true trial		-1			-2	-2		-1	-2		-1			-2	-2	

Fig. 2 Experimental procedure for both Tasks. **a** Experimental design for the explicit Belief and explicit Control SRT tasks, with Blocks numbered 1 to 30 on the first data row. In each block, there were 32 trials to which the participants had to respond: Standard (S) blocks with two repetitions of an embedded 16-trial Standard sequence; Random Orientation (RO) blocks with a pseudo-random Orientation sequence; or Total Random (TR) blocks with random sequences of Protagonists and Orientations. The RO and TR blocks were presented in two orders by switching the order of the RO and TR blocks (i.e., Order 1 & 2), counterbalanced between participants. **b** Standard sequence in the Belief SRT task. M = male (Papa Smurf); Fe = female (Smurfette); T = true; Fa = false. In the Control SRT task, the sequence is identical and stimuli were replaced by shapes and colors, as depicted in Fig. 1 (i.e., Male True = Blue Square; Male False = Orange Square, Female True = Green Circle, Female False = Black Circle). See Supplementary Table S1 for sequences in Random Blocks

except for the added explicit information of the existence of a sequence.

Experimental procedure

The experimental procedure was identical in both the explicit Belief and the explicit Control SRT tasks. Responses were made with the middle or index finger (i.e., 1 or 2 flowers respectively) of participants' left hand and were collected via a magnet compatible two-button response box. Responses were self-paced, and all stimuli remained on screen for 3,000 ms until a response was given. In case of a wrong response, or when no response was given after 3,000 ms, the word "Error" appeared for 750 ms on the screen, and the next trial began. The response-stimulus interval was set at 400 ms (Coomans et al., 2011). After each block, participants received feedback about their average reaction time (RT) and error rate and were encouraged to make less than 5% errors. For each block, a "Begin of block" and "End of block" message was presented for 4 s and 2 s respectively. Participants got a break of 15 s after every two blocks.

After a practice phase of 2 blocks of 24 trials (with a different sequence), the main task began. The current task consisted of 30 blocks with 32 trials each and was divided by the Training and Test phases (Fig. 2a).

In the initial Training phase, the Standard sequence was repeated throughout five blocks (Standard block). This Standard sequence consisted of 16-trials of Protagonist (Smurfs/shapes) and Orientation (beliefs/colors; Fig. 2b) and was repeated two times per block. Note that for a clear presentation, the shapes and color variation also are termed Protagonist and Orientation in the following text.

In the following Test phase, each Standard block, identical to those at Training phase, was followed by two types of Random blocks. First, in a Total Random block, Protagonist and Orientation were totally randomized with the limitation of at most two subsequent trials of the same Orientation type, consistent with the Standard blocks. Second, in a Random Orientation block, Orientation was changed into a different pseudo-random sequence while Protagonist remained identical as in the Standard blocks (Supplementary Table S1). The last block at the end of the whole task was always a Standard block. As Fig. 2b shows, there is a fixed sequence of the protagonists. However, because the previous behavioral study (Ma, Heleven, et al., 2021a) did not reveal sequence learning about protagonists, this sequence was not used to test random violations in the current or the previous neuroimaging study (Ma, Pu, et al., 2021b).

After scanning, participants were asked to reproduce, as accurately as possible, the "order in which Papa Smurf and Smurfette appeared" and whether they "could or could not see who gave the flowers" in the Belief SRT task; and the "order in which squares and circles appeared" for each combination of shape and color in the Control SRT task.

Imaging procedure and preprocessing

Images were collected with a Siemens Magnetom Prisma fit 3T scanner system (Siemens Medical Systems, Erlangen, Germany) using a 64-channel radiofrequency head coil. Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Stimulus presentation was controlled by E-Prime 2.0 (www.pstnet.com/eprime; Psychology Software Tools) running under Windows XP. Participants were placed head first and supine in the scanner bore and were instructed not to move their heads to avoid motion artifacts. Foam cushions were placed within the head coil to minimize head movements. First, high-resolution anatomical images were acquired using a T1-weighted 3D MPRAGE sequence [TR = 2,250 ms, TE = 4.18 ms, TI = 900 ms, FOV = 256 mm, flip angle = 9°, voxel size = 1 × 1 × 1 mm]. Second, a field map was calculated to correct for inhomogeneities in the magnetic field (Cusack & Papadakis, 2002). Third, whole-brain functional images were collected in a single run using a T2*-weighted gradient multiband echo sequence, sensitive to BOLD contrast (TR = 1,000 ms, TE = 31.0 ms, FOV = 210 mm, flip angle = 52°, slice thickness = 2.5 mm, distance factor = 0%, voxel size = 2.5 × 2.5 × 2.5 mm, 56 axial slices, acceleration factor GRAPPA = 4).

SPM12 (Wellcome Department of Cognitive Neurology, London, UK) was used to process and analyze the fMRI data. To remove sources of noise and artifact, data was preprocessed. Inhomogeneities in the magnetic field were corrected using the field map (Cusack & Papadakis, 2002). Functional data were corrected for differences in acquisition time between slices for each whole-brain volume, realigned to correct for head movement, and co-registered with each participant's anatomical data. Then, the functional data were transformed into a standard anatomical space (2-mm isotropic voxels) based on the ICBM152 brain template (Montreal Neurological Institute). Normalized data were then spatially smoothed (6-mm full-width at half-maximum, FWHM) using a Gaussian Kernel. Finally, using the Artifact Detection Tool (ART; http://web.mit.edu/swg/art/art.pdf; http://www.nitrc.org/projects/artifact_detect), the preprocessed data were examined for excessive motion artifacts and for correlations between motion and experimental design, and between global mean signal and experimental design. Outliers were identified in the temporal differences series by assessing between-scan differences (Z-threshold: 3.0 mm, scan to scan movement threshold: 0.5 mm; rotation threshold: 0.02 radians). These outliers were omitted from the analysis by including a single regressor for each outlier. A default high-pass filter was used for 128 s, and

serial correlations were accounted for by the default auto-regressive AR (1) model.

Statistical analysis of neuroimaging data

In all statistical analyses (and preprocessing) that are described below, we used exactly the same parameters as in Ma, Pu, et al. (2021b) to facilitate comparability between studies.

The statistical analyses were performed using the general linear model of SPM12 (Wellcome Department of Cognitive Neurology, London, UK). At the first (single participant) level, an event-related design for measuring transient activity across trials was modeled by entering separate regressors for the trials of interest: two regressors for the trials in the Standard blocks at the Training and Test phases (i.e., Standard block at Training, Standard block at Test), two regressors for the trials in the Total Random blocks and the trials in the Random Orientation blocks at the Test phase (i.e., Total Random at Test, Random Orientation at Test), and two additional regressors of no interest for pauses and error trials. This last regressor involved incorrect trials as well as one trial after each incorrect trial, because these latter trials may be affected by error processing on the prior trial.

Sequence learning effects within explicit Belief and explicit Control SRT tasks

At the second (group) level, we defined the following three contrasts related to explicit sequence learning effects:

1) *General learning:* brain activations during initial learning of the Standard sequence are tested by the contrast: Standard block at Training > Standard block at Test.
2) *Maintenance of learning:* brain activations during late learning in a context of sequence violations (in the Test phase) are tested by the contrast: Standard block at Test > Standard block at Training. Note that this contrast does not show mere late phase of sequence learning, as it also involves reinstating the learned Standard sequence after Random sequences.
3) *Detecting violations:* brain activations for detecting sequence violations among the learned Standard sequence in the Test phase, are tested by two contrasts: Total Random block at Test > Standard block at Test and Random Orientation block at Test > Standard block at Test.

Because we hypothesized that the cerebellar Crus and TPJ are preferentially engaged in the explicit Belief task, but less so in the explicit Control SRT task, we ran these three contrasts for each task separately. We conducted a within-participant one-way analysis of variance (ANOVA) and

defined t-contrasts between four regressors of interests (i.e., Standard block at Training, Standard block at Test, Total Random block at Test, and Random Orientation block at Test). Clusters of activation were first defined by a cluster-forming threshold of $p < 0.001$ (uncorrected) with a minimum extent of 10 voxels (Flandin & Friston, 2019), and we further restricted significant clusters using a cluster-wise significance level of $p < 0.05$, family-wise error (FWE) correction for multiple comparisons.

To further explore and visualize the neural time course of the posterior cerebellar Crus II and TPJ throughout the explicit Belief SRT task, we ran another model with 32 regressors (i.e., for 30 blocks and two additional regressors of no interest for errors and pause trials). Percent signal change data of the peak coordinates of these areas at each of the above three contrasts were extracted using the MarsBar toolbox (http://marsbar.sourceforge.net), using a sphere with a radius of 10 mm. For each participant, percent signal change data were computed for each block. In addition, a Spearman correlation analysis between percent signal change and RTs was computed. To be exhaustive, we also ran this analysis in the explicit Control SRT task if any cerebellar and TPJ activations were found.

Sequence learning effects across tasks

Sequence learning effects across explicit Belief and explicit Control SRT tasks To test preferential activation of the cerebellar Crus and TPJ in the explicit Belief SRT task as opposed to the explicit Control SRT task, we modeled eight covariates with Task (explicit Belief vs. explicit Control) as a between-participant factor orthogonal to the same four regressors of interests as before as a within-participant factor (i.e. Standard block at Training, Standard block at Test, Total Random block at Test, and Random Orientation block at Test).

We first tested simple contrasts between the two tasks (i.e., explicit Belief > explicit Control SRT tasks) for each block type (Standard and Random) and at all phases (Training and Test). Importantly, to test our hypothesis that the sequence learning effects are asymmetric directly (i.e., engaged more in the explicit Belief SRT task than the explicit Control SRT task), a series of asymmetric interaction effects was defined for each of the sequence learning effects, also known as *spreading interactions*. For example, the spreading interaction for *general learning* was: explicit Belief Standard block at Training > explicit Belief Standard block at Test = [explicit Control Standard block at Training = explicit Control Standard block at Test], or expressed in weights: 3 −1 −1 −1. As a comparison, we also ran reverse *spreading interactions* with an asymmetric effect in the explicit Control SRT task.

Sequence learning effects across implicit and explicit Belief SRT tasks For the comparison between the present explicit Belief SRT task and the previously published implicit Belief SRT task by Ma, Pu, et al. (2021b), we modeled eight covariates with Task (Implicit versus Explicit Belief SRT task), as additional between-participant factor orthogonal to the same four regressors of interests as before, as a within-participant factor (i.e., Standard block at Training, Standard block at Test, Total Random block at Test, and Random Orientation block at Test).

As before, we tested simple contrasts (i.e., implicit Belief > explicit Belief SRT tasks) as well as the *spreading interaction*. For example, to test that the cerebellar Crus was engaged more in the implicit than explicit Belief SRT task, the spreading interaction for *general learning* was: implicit Belief Standard block at Training > implicit Belief Standard block at Test = [explicit Belief Standard block at Training = explicit Belief Standard block at Test =]. We also ran reverse interactions with stronger asymmetric effects in the explicit Belief SRT task.

All between task comparisons outlined above were conducted using the Sandwich Estimator toolbox (SwE; Guillaume et al., 2014; http://www.nisox.org/Software/SwE/). SwE uses a marginal model to analyze repeated measurements between tasks, taking into account correlations because of repeated measurements, unexplained variations across participants, unbalanced study designs of the variable number of scans, and corrected degrees of freedom. We used the following default SwE options (see http://www.nisox.org/Software/SwE/man): a modified SwE which assumes that participants in each task share a common covariance matrix, repeated measurements in each within-factor regressor, small-sample adjustment = type C2, and degrees of freedom = approximation III, Non-Parametric Wild Bootstrap = No. Note that we used a parametric approach for thresholding rather than non-parametric bootstrap because the parametric approaches are statistically more efficient, reproducible and computationally more efficient than their nonparametric counterpart (Flandin & Friston, 2019). The SwE contrasts were analyzed using a cluster-forming threshold of $p < 0.005$ with minimum cluster extent of 50 voxels, followed by a voxel level significance of $p < 0.05$, using false-discovery rate (FDR) correction for multiple comparisons, which is only option in SwE for parametric thresholds. As before, we begin reporting the results of the ROI analysis and followed by the remaining results of the whole-brain analysis.

ROI analyses

Based on previous fMRI and meta-analyses, we defined a number of *a priori* Regions of Interest (ROI) of the posterior cerebellar Crus I & II and the cortical TPJ:

1) Meta-analyses showed significant activations of the bilateral Crus II (± 24 -76 -40) during social reasoning (see also Guell et al., 2018; Van Overwalle, Ma, & Heleven, 2020a). The bilateral Crus II are located within the mentalizing network demarcated by Buckner et al. (2011; Fig. 3a and b). We also identified left Crus I (-40 -70 -40) as a ROI, which showed activations in previous fMRI studies (Heleven et al., 2019; Ma, Pu, et al., 2021b); note that this area is located somewhat more peripherally in the mentalizing network and closer to the executive network (Buckner et al., 2011; Fig. 3c).

2) Meta-analyses showed significant activations of the bilateral TPJ ± 50 -55 25 in understanding people's beliefs, intentions, and personality traits (Van Overwalle, 2009; Van Overwalle & Baetens, 2009).

The *a priori* ROIs were tested with a small volume correction (SVC) using spheres with radius = 10 mm centered around the nearest 2 mm of the coordinates listed above (Calvo-Merino et al., 2005; Debas et al., 2010), using the
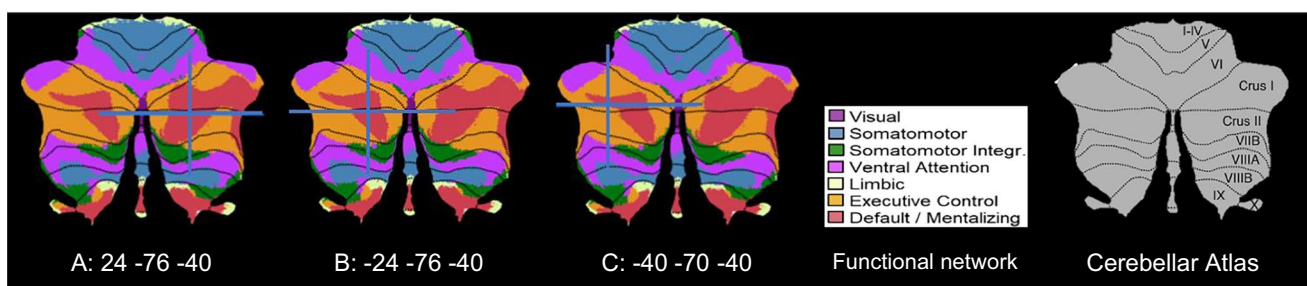
Fig. 3 *A priori* cerebellar ROIs drawn on flat maps (Diedrichsen & Zotow, 2015, http://www.diedrichsenlab.org/imaging/AtlasViewer/viewer.html) showing the functional networks by Buckner et al. (2011), with ROI centers indicated by blue crosshairs. Our bilateral Crus II ROIs are clearly located within the mentalizing network denoted by red color, and Crus I is located close to the border of the mentalizing network (Buckner et al., 2011)

same thresholds as for the whole-brain analysis (except that the minimum cluster extent was always set to 10 voxels).

Statistical analysis of behavioral data

For behavioral data, responses during and immediately after an error were excluded for computing mean reaction times (RTs). Mean error rates and mean correct RTs were computed for every block.

To assess learning of the Standard sequence, we used a mixed ANOVA with the all Standard blocks at Training and Test as within-participant factors and Task (explicit Belief versus explicit Control) as a between-participant factor. To test the pattern of RTs for detecting violations, we also used a mixed ANOVA with Blocks at Test and Block Type (Standard, Total Random and Random Orientation blocks) as within-participant factors and Task (explicit Belief vs. explicit Control) as a between-participant factor. If participants learned the Standard sequence, their RTs should decrease across the Standard blocks, and increase again during the Random blocks.

For the statistical results, the Greenhouse-Geisser correction is reported when the sphericity assumption was violated. T-tests were applied when ANOVA indicated significant differences. The level of significance was set to 0.05, and two-tailed tests were applied.

Power analyses for the neuroimaging and behavioral data

To identify the minimum size of the effect that can be reliably detected, we applied a sensitivity power analyses on the current experiment by using G*power 3.1.9.4 (Faul et al., 2009). This method also was used in a previous neuroimaging study (Gao et al., 2019).

For a sensitivity power analysis of the neuroimaging data, criteria were set for a t-test with an alpha significance criterion of 0.05, standard power criterion of 95% and a sample of 22 for each of the tasks or a sample of 44 across two tasks, G*power resulted in an effect size of *Cohen's d =* 0.81 (which equal to $t = 2.08$) for sequence learning effects within explicit Belief and explicit Control SRT tasks, or an effect size of *Cohen's d =* 1.11 (which equal to $z = 3.2$) for sequence learning effects across the two tasks (note that G*power outputs *Cohen's d* and *t* values, and these values were further converted to *z*-value using the current sample size by http://psychometrica.de/effect_size.html). As shown in Tables 1 and 2 and Tables S2-S3, the *t*- or *z*-values from the peak activations met these requirements.

For the behavioral data, criteria were set for a mixed ANOVA with an alpha significance criterion of 0.05, and standard power criterion of 95%. The number of groups was set to 2 (i.e., two tasks), and the number of repeated

measurements 14 (i.e., 14 Standard Blocks in the Training and Test Phases), or 24 (i.e., Standard, Total Random and Random Orientation blocks in the Test Phase). G*power resulted in the minimum effect of $\eta^2 = 0.02$ for learning of the Standard sequence and $\eta^2 = 0.01$ for RT patterns of detecting violations. Our results showed that the significant findings in our experiment met these effect size requirements.

Overall, the sensitivity power analyses indicated that our sample provided well-powered and reliable behavioral and neuroimaging results in the explicit Belief and explicit Control SRT tasks. Also, our sample size was comparable to previous neuroimaging studies on sequence learning as reported in meta-analyses (an average sample size of 14; Hardwick et al., 2013; Janacsek et al., 2020).

Results

Behavioral results

We analyzed participants' error rates and mean reaction times (RTs). The error rate was low for the explicit Belief (5.3%) and explicit Control (5.9%) SRT tasks over all blocks. The participants learned the Standard sequence, as demonstrated by significant faster RTs during the Standard blocks at Training and Test phase, and significant slower RTs during the Random compared to the Standard blocks at the Test phase. These results were supported by the following statistical analyses.

To test learning of the Standard sequence, we first applied a mixed repeated-measures ANOVA with 14 Standard blocks at Training and Test phases as within-participant factor and Tasks (explicit Belief vs. explicit Control) as between-participant factor. A significant linear trend revealed faster RTs across Standard blocks ($F_{(1, 42)} = 205.49$, $p < 0.001$, $\eta^2 = 0.83$), with marginal significant differences between the explicit Belief and explicit Control SRT tasks, indicating slightly faster RTs in the explicit Belief SRT task (Bonferroni post-hoc test: Mean RTs Difference (MD) $_{Belief-Control}$ = 58.16 ms, $F_{(1, 42)} = 3.23$, $p = 0.08$, $\eta^2 = 0.07$), and no interaction ($p = 0.9$, Fig. 4a).

Second, we ran a mixed repeated-measures ANOVA at the Test phase with Block Type (Standard, Total Random and Random Orientation blocks) and Blocks at Test phase as within-participant factors and Task (explicit Belief versus explicit Control) as a between-participant factor (Fig. 4a). A significant main effect of Block Type confirmed that participants reacted slower in the Total Random and Random Orientation blocks than the Standard blocks ($F_{(1.76, 73.81)} = 9.77$, $p < 0.001$, $\eta^2 = 0.19$, MD $_{Total\ Random\ –\ Standard\ block}$ = 24 ms, $p < 0.01$; MD $_{Random\ Orientation\ –\ Standard\ block}$ = 22 ms, $p < 0.01$). A significant linear trend in the Blocks at Test

Table 1 Whole-brain and ROI activations of spreading interaction between explicit Belief and explicit Control SRT tasks by SwE analysis

Contrasts and Anatomical Label	MNI coordinate					Contrasts and anatomical label	MNI coordinate				
	x	y	z	Voxels	max z		x	y	z	Voxels	max z
General learning											
explicit Belief Standard block at Training > explicit Belief Standard block at Test [=explicit Control Standard block at Training = explicit Control Standard block at Test]						explicit Control Standard block at Training > explicit Control Standard block at Test [=explicit Belief Standard block at Training = explicit Belief Standard block at Test]					
ROI analysis						ROI analysis					
R Temporoparietal Junction	44	-58	20	193	4.470**	--					
	50	-60	18		4.152**						
	48	-48	28		3.856**						
	50	-46	20		3.621**						
	56	-46	26		3.152**						
Whole-brain analysis						Whole-brain analysis					
R Lingual Gyrus	4	-84	-10	170	3.929**	R Superior Occipital Gyrus	22	-94	20	441	4.814
L Middle Occipital Gyrus	-46	-80	12	1649	4.877***	Area hOc1 [V1]	-16	-86	2	158	4.221
L Fusiform Gyrus	-38	-58	-14		4.754***	Area hOc1[V1]	-30	-58	0	230	3.867~
L Middle Occipital Gyrus	-46	-72	10		4.898***						
R Middle Temporal Gyrus	52	-64	12	4002	6.222***						
R Fusiform Gyrus	36	-58	-16		6.076***						
R Middle Occipital Gyrus	46	-76	0		5.447***						
R Precuneus	6	-48	52	1110	4.044**						
R Precuneus	8	-64	58		3.948**						
L Precuneus	-12	-58	52		3.960**						
L Supra Marginal Gyrus	-64	-22	36	58	3.242*						
R Middle Frontal Gyrus	36	-2	54	633	4.227**						
R Middle Frontal Gyrus	42	6	38		3.529*						
R Middle Frontal Gyrus	48	10	48		3.989**						
L Precentral Gyrus	-30	0	58	215	3.565*						
R Inferior Frontal Gyrus (p. Opercularis)	54	10	26	55	3.199*						
R Posterior-Medial Frontal	6	16	46	163	3.530*						
Maintenance of learning											
explicit Belief Standard block at Test > explicit Belief Standard block at Training [=explicit Control Standard block at Test = explicit Control Standard block at Training]						explicit Control Standard block at Test > explicit Control Standard block at Training [= explicit Belief Standard block at Test = explicit Belief Standard block at Training]					
ROI analysis						ROI analysis					
R Cerebellum (Crus II)	18	-80	-42	246	4.878***	R Cerebellum (Crus II)	32	-80	-36	74	3.254*
	26	-76	-34		4.547***		30	-74	-38		3.209*
	20	-76	-34		4.441***		18	-82	-36		2.932*
L Cerebellum (Crus II)	-14	-76	-40	238	4.193**		22	-78	-34		2.742*
	-28	-78	-36		3.608**						
	-20	-78	-44		3.548**						
R Temporoparietal Junction	50	-46	20	103	4.030**						
	52	-58	18		2.950*						
	42	-54	20		2.699*						
	46	-56	18		2.676*						
L Temporoparietal Junction	-58	-58	22	125	3.632*						
	-60	-54	26		3.615*						
	-56	-62	26		3.433*						
	-50	-62	20		3.108*						
	-50	-48	18		2.989*						
	-46	-48	22		2.858*						
	-50	-52	26		2.795*						
Whole-brain analysis						Whole-brain analysis					
R Cerebellum (Crus II)	18	-80	-42	439	4.878*	--					

Table 1 (continued)

Contrasts and Anatomical Label	MNI coordinate					Contrasts and anatomical label	MNI coordinate				
	x	y	z	Voxels	max z		x	y	z	Voxels	max z
R Cerebellum (Crus II)	18	-86	-32		4.585*						
R Cerebellum (Crus II)	26	-76	-34		4.547*						
L Cerebellum (Crus II)	-14	-76	-40	380	4.193*	--					
L Cerebellum (Crus II)	-28	-78	-36		3.608*						
L Cerebellum (Crus II)	-16	-86	-34		3.020~						
L Calcarine Gyrus	0	-88	-6	173	3.881*	R Superior Occipital Gyrus	20	-94	18	505	5.096*
L Middle Temporal Gyrus	-50	-72	18	1770	4.462*	R Superior Occipital Gyrus	26	-90	24		4.925*
L Middle Temporal Gyrus	-60	-58	16		4.075*	R Calcarine Gyrus	14	-90	12		4.514*
L Middle Temporal Gyrus	-50	-38	4		4.330*	L Superior Occipital Gyrus	-10	-94	4	330	4.836*
R Middle Temporal Gyrus	50	-66	14	771	4.633*	L Middle Temporal Gyrus	-54	-28	-2	54	3.835*
R Middle Temporal Gyrus	58	-54	14		3.597*	R Superior Temporal Gyrus	60	-12	-2	98	3.390~
R Superior Temporal Gyrus	50	-44	20		4.264*	L Pallidum	-22	-4	0	294	4.200*
L Fusiform Gyrus	-40	-58	-18	137	3.637*	R Pallidum	18	6	-4	341	4.288*
R Precuneus	4	-54	14	104	3.280*	L Temporal Pole	-52	10	-16	384	4.364*
R Fusiform Gyrus	42	-48	-16	223	4.485*						
L Midcingulate Cortex	-6	-46	36	254	4.904*						
L Paracentral Lobule	-4	-32	74	159	4.043*						
L Postcentral Gyrus	-60	-12	40	119	3.486*						
R Medial Temporal Pole	56	2	-18	306	4.204*						
L Superior Medial Gyrus	-10	42	42	534	4.234*						
L Superior Medial Gyrus	-4	48	44		3.897*						
L Superior Medial Gyrus	-8	50	34		4.146*						

Detecting violations

explicit Belief Total Random & Random Orientation block at Test > explicit Belief Standard block at Test [=explicit Control Standard block at Test = explicit Control Total Random & Random Orientation block at Test]

explicit Control Total Random & Random Orientation block at Test > explicit Control Standard block at Test [=explicit Belief Standard block at Test = explicit Belief Total Random & Random Orientation block at Test]

ROI analysis

ROI analysis

	x	y	z	Voxels	max z		x	y	z	Voxels	max z
L Cerebellum (Crus II)	-14	-76	-40	66	4.782***	--					
R Temporoparietal Junction	48	-48	28	373	4.958***						
	56	-46	26		4.546***						
	50	-46	20		4.237***						
	50	-60	18		4.206***						
	56	-54	18		3.490***						
L Temporoparietal Junction	-50	-62	20	34	3.670*						
	-54	-62	22		3.371*						

Whole-brain analysis

Whole-brain analysis

	x	y	z	Voxels	max z		x	y	z	Voxels	max z
L Cerebellum (Crus II)	-14	-74	-40	165	4.891***	--					
R Middle Temporal Gyrus	52	-66	14	5618	5.835***	R Superior Occipital Gyrus	22	-94	18	520	5.359**
R Lingual Gyrus	6	-84	-10		5.062***	R Superior Occipital Gyrus	26	-90	24		4.804**
R Angular Gyrus	48	-48	28		4.958***	R Calcarine Gyrus	14	-90	10		4.410*
R Precuneus	4	-46	42	671	4.159**	L Superior Occipital Gyrus	-12	-92	4	272	4.862**
R Precuneus	4	-46	52		4.064**						
L Inferior Parietal Lobule	-22	-50	52		3.305*						
L Postcentral Gyrus	-60	-12	40	83	3.577*						
R Superior Frontal Gyrus	32	-10	44	93	3.599*						
L Precentral Gyrus	-36	-4	44	96	3.220*						
R Middle Frontal Gyrus	48	12	46	79	3.184*						

Notes: Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. Whole-brain and ROI (small volume correction) analysis in SwE, using a cluster-forming threshold of $p < 0.005$ (uncorrected) with minimum cluster extent of 50 voxels (for the ROI 10 voxels), followed by a peak-wise threshold of $p < 0.05$, FDR corrected. Only the highest peaks of each cluster are shown, except for the cerebellum, TPJ, and related ROIs, as well as for large clusters > 500 voxels. L = left, R = right. ROIs for the posterior cerebellum are centered around coordinates ±24, -76, -40 (Crus II); and -40, -70, -40 (Crus I) and for the TPJ around coordinates ±50, -55, 25, all spheres with radius 10 mm. ROIs that are not mentioned are not significant

~$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (peak FDR corrected)

Table 2 Whole-brain and ROI activations of spreading interaction between the implicit and explicit Belief SRT tasks by SwE analysis

Contrasts and anatomical label	MNI coordinate					Contrasts and anatomical label	MNI coordinate				
	x	y	z	Voxels	max z		x	y	z	Voxels	max z
General learning											
implicit Belief Standard block at Training > implicit Belief Standard block at Test [= explicit Belief Standard block at Training = explicit Belief Standard block at Test]						explicit Belief Standard block at Training > explicit Belief Standard block at Test [= implicit Belief Standard block at Training = implicit Belief Standard block at Test]					
ROI analysis						ROI analysis					
L Cerebellum (Crus I)	-36	-64	-34	25	3.08~	--					
	-44	-68	-32		2.71~						
Whole-brain analysis						Whole-brain analysis					
--						--					
Maintenance of learning											
implicit Belief Standard block at Test > implicit Belief Standard block at Training [= explicit Belief Standard block at Test = explicit Belief Standard block at Training]						explicit Belief Standard block at Test > explicit Belief Standard block at Training [=implicit Belief Standard block at Test = implicit Belief Standard block at Training]					
ROI analysis						ROI analysis					
R Cerebellum (Crus II)	24	-78	-34	146	5.00***	R Cerebellum (Crus II)	20	-82	-42	150	3.65**
							30	-82	-38		3.36**
							18	-78	-34		3.10**
							32	-76	-34		2.99**
							20	-84	-36		2.73**
L Cerebellum (Crus II)	-24	-80	-36	81	3.34*						
Whole-brain analysis						Whole-brain analysis					
R Cerebellum (Crus I)	24	-78	-34	207	5.00*	--					
L Cerebellum (Crus II)	-24	-80	-36	86	3.34~						
L Cerebellum (IX)	-8	-44	-38	43	3.52~						
L Middle Temporal Gyrus	-50	-12	-10	1316	4.62*						
L Middle Temporal Gyrus	-44	-18	-12		4.47*						
L Superior Temporal Gyrus	-46	-2	-14		4.44*						
R Superior Temporal Gyrus	60	-8	-2	167	3.67~						
L Putamen	-16	8	-6	191	4.12*						
R Caudate Nucleus	12	14	-6	239	4.29*						
Detecting violations											
implicit Belief Total Random & Random Orientation block at Test > implicit Belief Standard block at Test [=explicit Belief Standard block at Test = explicit Belief Total Random & Random Orientation block at Test]						explicit Belief Total Random & Random Orientation block at Test > explicit Belief Standard block at Test [= implicit Belief Standard block at Test = implicit Belief Total Random & Random Orientation block at Test]					
ROI analysis						ROI analysis					
--						R Temporoparietal Junction	56	-46	26	37	3.27*
							54	-46	30		3.17*
							48	-50	32		2.93*
							54	-46	22		2.66*
						R Temporoparietal Junction	48	-62	26	32	3.11*
							52	-62	28		3.06*
Whole-brain analysis						Whole-brain analysis					
--						--					

Notes: Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. Whole-brain and ROI (small volume correction) analysis in SwE, using a cluster-forming threshold of $p < 0.005$ (uncorrected) with minimum cluster extent of 50 voxels (for the ROI and cerebellum 10 voxels), followed by a peak-wise threshold of $p < 0.05$, FDR corrected. Only the highest peaks of each cluster are shown, except for the cerebellum, TPJ, and related ROIs, as well as for large clusters > 500 voxels. L = left, R = right. ROIs for the posterior cerebellum are centered around coordinates ±24, -76, -40 (Crus II); and -40, -70, -40 (Crus I) and for the TPJ around coordinates ±50, -55, 25, all spheres with radius 10 mm. ROIs that are not mentioned are not significant

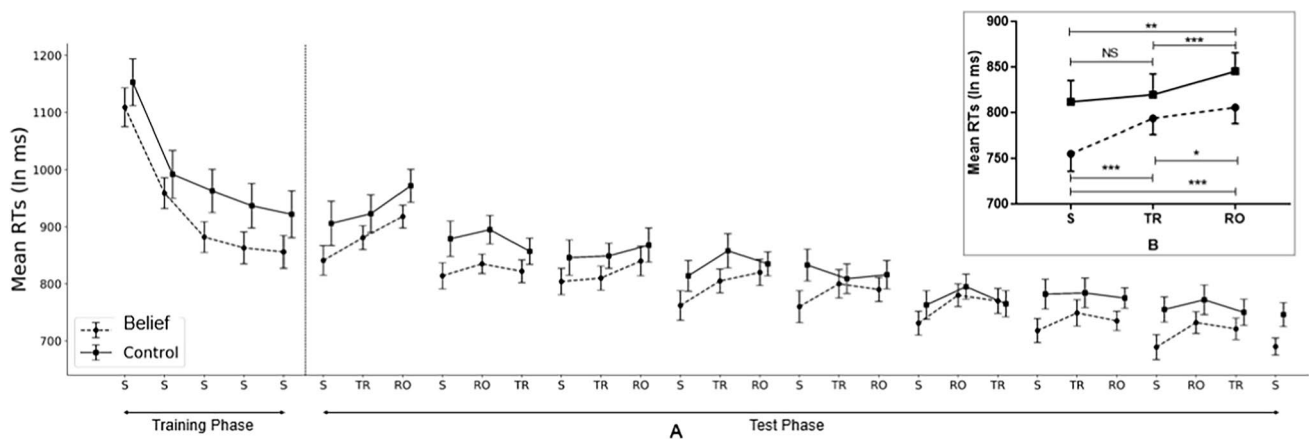$\sim p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ (peak FDR corrected)

Fig. 4 **a** Behavioral performance demonstrated by mean RTs at the explicit Belief SRT task (dashed lines) and explicit Control SRT task (full lines) for each block. **b** Collapsed RTs for Standard and Random blocks at the Test phase. Error bars are within-tasks standard error of the mean across participants. S = Standard Block, TR = Total Random block, RO = Random Orientation block. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. NS = not significant

suggests that there was a decrease in RTs over the Test phase ($F_{(1, 42)} = 185.59$, $p < 0.001$, $\eta^2 = 0.82$). There was no main effect of Task ($p > 0.1$).

There was an interaction between Block Type and Task ($F_{(1.76, 73.81)} = 3.23$, $p = 0.05$, $\eta^2 = 0.07$). Simple effect analysis suggested that the pattern of slower responses in the two random blocks was strongest in the explicit Belief SRT task ($F_{(2, 41)} = 8.68$, $p = 0.001$, $\eta^2 = 0.30$). Closer inspection using paired t-tests (Fig. 4b) revealed reliably slower RTs for the Total Random blocks than the Standard blocks at Test phase only in the explicit Belief SRT task (Belief: $t_{(21)} = 4.53$, $p < 0.001$, *Cohen's d* = 0.99; Control: $t_{(21)} = 0.94$, $p = 0.36$, *Cohen's d* = 0.14), and slower RTs for the Random Orientation blocks than the Standard blocks at Test phase for both explicit SRT tasks (Belief: $t_{(21)} = 6.36$, $p < 0.001$, *Cohen's d* = 1.27; Control: $t_{(21)} = 3.49$, $p = 0.002$, *Cohen's d* = 0.71), and slower RTs for the Random Orientation blocks than the Total Random blocks for both tasks (Belief: $t_{(21)} = 2.08$, $p = 0.05$, *Cohen's d* = 0.14; Control: $t_{(21)} = 4.08$, $p < 0.001$, *Cohen's d* = 0.25) .

Once outside the scanner, participants were asked to reproduce the sequence as accurately as possible. The mean length of the longest correct sequence reproduction was 4.45/16 in the explicit Belief SRT task, and 2.95/16 in the explicit Control SRT task.

Neuroimaging results: Comparisons within the Explicit SRT task (cf. Hypothesis 1)

We mainly report on three sequence learning effects of interest: *general learning* (Standard block at Training > Standard block at Test), *maintenance of learning* (Standard block at Test > Standard block at Training) and *detecting violations* (Random block at Test > Standard block at Test).

In reporting the results, we begin with our *a priori* ROIs, followed by the remaining exploratory results of the whole-brain analysis.

Sequence learning effects within explicit Belief and explicit Control SRT tasks

We first tested the three sequence learning effects on the activation of Crus II and TPJ within the explicit Belief SRT and explicit Control SRT tasks separately. For the explicit Belief SRT task, the ROI results showed that the cerebellar Curs I and II were activated during *maintenance of learning* (Fig. 5a), and the TPJ was activated during *general learning* (Fig. 5b) and *detecting violations* (Fig. 5c and d). For the explicit Control SRT task, none of our TPJ ROIs were activated; however, there was an unexpected bilateral activation of the cerebellar Crus II during *maintenance of learning*. See Supplementary Table S2 for detailed ROI and whole-brain results.

To further explore potential relationships between the behavioral level (i.e., RTs) and the neurological level (i.e., brain activations) during sequence learning, we further extracted the percent signal change of the peak activations in the cerebellar Crus II and TPJ for each of the above contrasts, and computed a Spearman correlation with RTs across all 30 blocks. These peak activations are shown in Fig. 5 for the bilateral cerebellar Crus II (MNI 26 −76 −36 & −26 −78 −34 from the *maintenance of learning* contrast in Fig. 5a) and for the right TPJ (MNI 44 −58 20 from the *general learning* contrast in Fig. 5b; MNI 52 −48 28 from the *detecting violations* in Orientations contrast in Fig. 5d). To compute the percent signal change, we used ROIs with a radius of 10 mm centered at these peak activations. The correlational analysis showed that there were significant negative correlations between RTs and the bilateral cerebellar
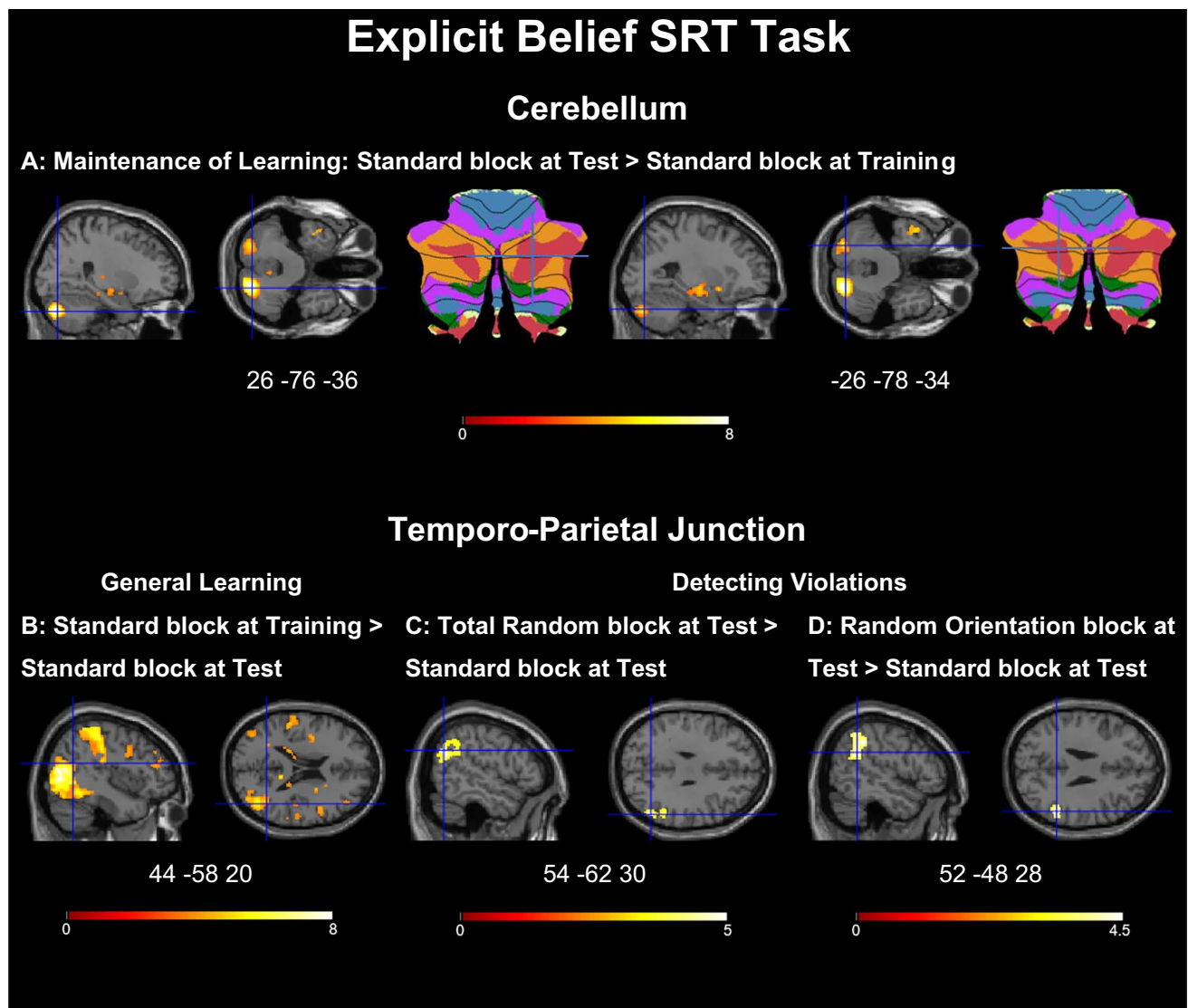
Fig. 5 Activation of the cerebellar Crus and TPJ during sequence learning effects within the explicit Belief SRT task, displayed at an uncorrected threshold of $p < 0.001$, with color bars denoting t values. **a** Sagittal and transverse views of whole-brain activation at the peak coordinates of the cerebellar Crus I & II, indicated by crosshairs dis- played on brain slices and on flatmaps(Diedrichsen & Zotow, 2015) depicting functional networks (Buckner et al., 2011). **b-d** Sagittal and transverse views of whole-brain activation at the peak coordinates of the TPJ, indicated by crosshairs. Note that not all visible clusters are significant after FWE correction

Crus II (right: $r = -0.75$, $p < 0.001$; left: $r = -0.40$, $p = 0.03$). Supplementary Figure S1 shows that these correla- tions reflect patterns of gradually faster RTs and higher Crus II activation over the course of the experiment. In addition, there were significant positive correlations between RTs and TPJ activation (from the *general learning* contrast: $r = 0.84$, $p < 0.001$; from *detecting violations* in Orientations contrast: $r = 0.35$, $p = 0.05$). Supplementary Figure S2 shows that these correlations reflect patterns of gradually faster RTs and lower TPJ activations for the standard sequence over the course of the experiment, and slower RTs and higher TPJ activations for violations.

To be exhaustive, we also extracted the percent signal change of the cerebellar peak activations revealed in the explicit Control SRT task. However, although the cerebel- lar Crus was activated during *maintenance of learning*, there were no significant correlations between the bilateral cer- ebellar Crus II activation and RTs (all $ps > 0.2$).

Simple contrasts between explicit Belief and explicit Control SRT tasks

To further test the preferential activations of Crus II and TPJ in the explicit Belief SRT task in comparison to the explicit Control SRT task, we conducted simple

contrasts between explicit Belief versus explicit Control for all Block Types (Standard and Random) at all Phases (Training and Test) using the SwE analysis. Overall, ROI analyses revealed stronger activation of the left posterior cerebellar Crus II in the Belief > Control contrast for all Block Types and all Phases, except the Standard block at Training (Fig. 6a). The ROI analyses also revealed stronger activation of the right TPJ in the explicit Belief > explicit Control contrast for all Block Types and all Phases, and stronger activation of the left TPJ in the same contrast in the Standard block at Test and Total Random at Test (Fig. 6b). The reverse explicit Control > explicit Belief comparisons did not reveal any activations of interest. We report the exploratory whole-brain results in Supplementary Table S3.

Spreading interactions across explicit Belief versus explicit Control SRT tasks

To test our hypothesis directly, we applied a series of *spreading interactions* that assume higher activation in the explicit Belief SRT task for the three learning effects and none in the explicit Control SRT task. Almost all the hypothesized stronger activations in the cerebellar Crus and the cortical TPJ were revealed in the explicit Belief SRT task compared with the explicit Control SRT task (Table 1).

First, for the spreading interaction of *general learning*, the ROI analysis only showed activation of the TPJ, but not of the cerebellar Crus I & II (Fig. 7a right). The whole-brain analysis further revealed activations in the lingual gyrus, middle occipital gyrus, middle/inferior temporal gyrus, precuneus, supramarginal gyrus, middle frontal gyrus, and posterior-medial frontal.

Second, for the spreading interaction of *maintenance of learning*, the ROI and whole-brain analyses showed bilateral
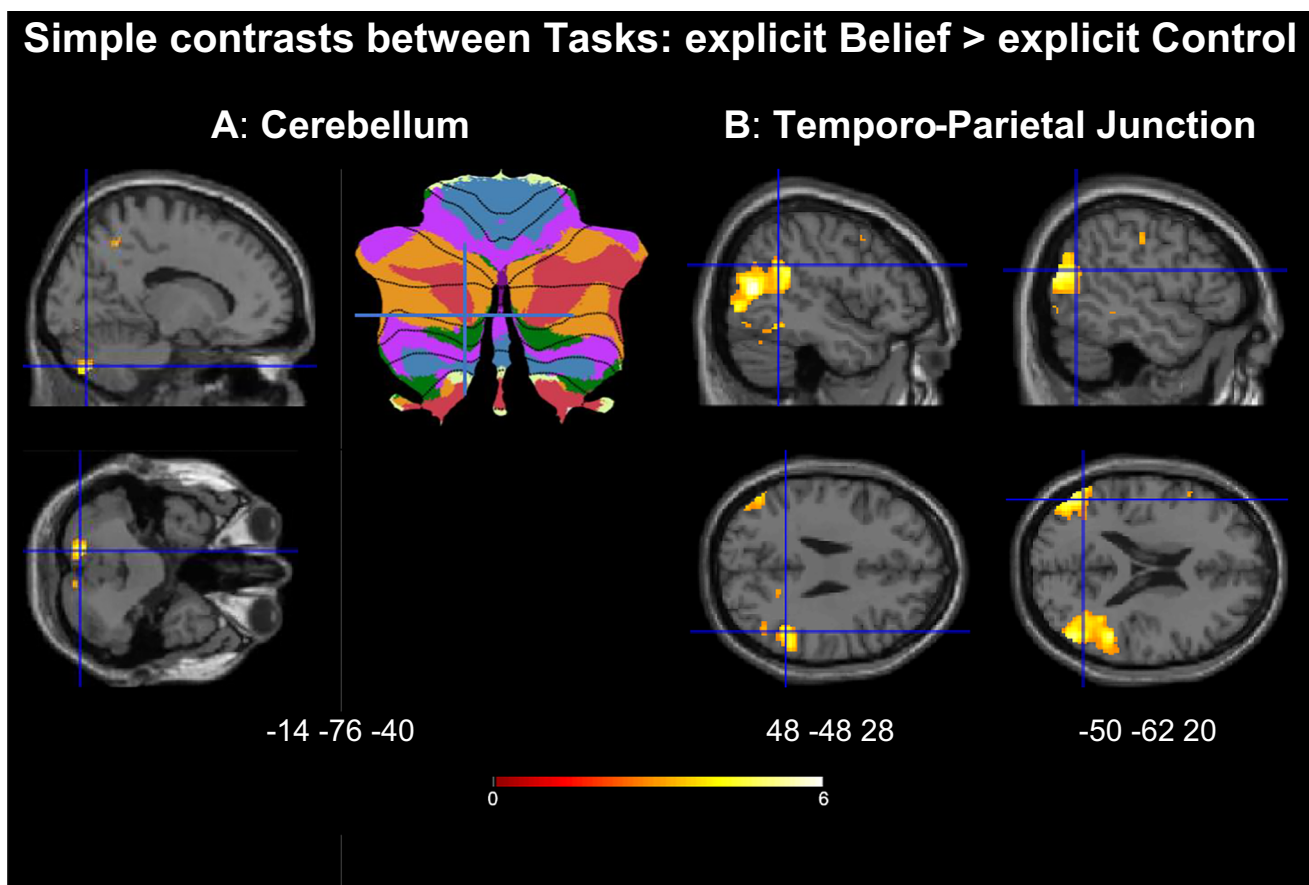

Fig. 6 Simple contrasts showing stronger cerebellar Crus II and the TPJ activation at the explicit Belief > explicit Control SRT task. Sagittal and transverse views of brain activation at the peak coordinates of cerebellar Crus II (**a**) and the TPJ (**b**) and, indicated by crosshairs, displayed at an uncorrected threshold of $p < 0.005$, with color bars denoting SwE *z* value. The peak cerebellar activation is also indicated with a blue crosshair on a flatmap (Diedrichsen & Zotow, 2015) displaying functional networks by Buckner et al. (2011). Activations of TPJ and Crus II are taken from the Total Random block: explicit Belief > explicit Control contrast which showed the largest cluster. Note that not all visible clusters are significant after FDR correction

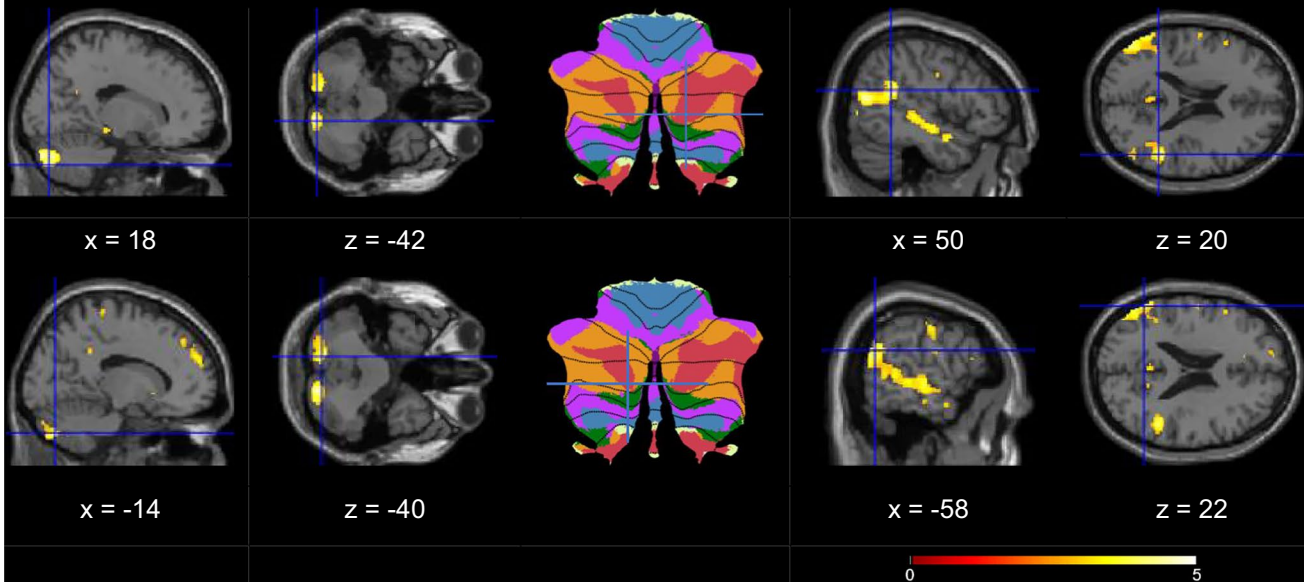**Spreading interactions: explicit Belief versus explicit Control SRT tasks**

Cerebellum   **Temporo-Parietal Junction**

A: explicit Belief Standard block at Training > explicit Belief Standard block at Test [= explicit Control Standard block at Training = explicit Control Standard block at Test]

x = 44   z = 20

B: explicit Belief Standard block at Test > explicit Belief Standard block at Training [=explicit Control Standard block at Test = explicit Control Standard block at Training]

x = 18   z = -42   x = 50   z = 20

x = -14   z = -40   x = -58   z = 22
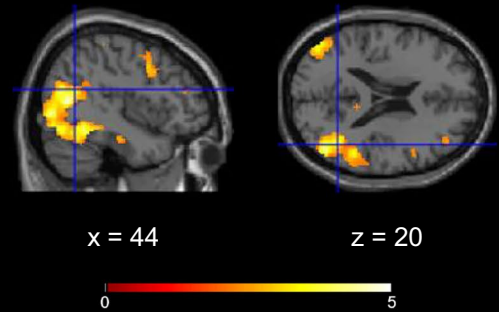
C: explicit Belief Random block at Test > explicit Belief Standard block at Test [=explicit Control Standard block at Test = explicit Control Random block at Test]
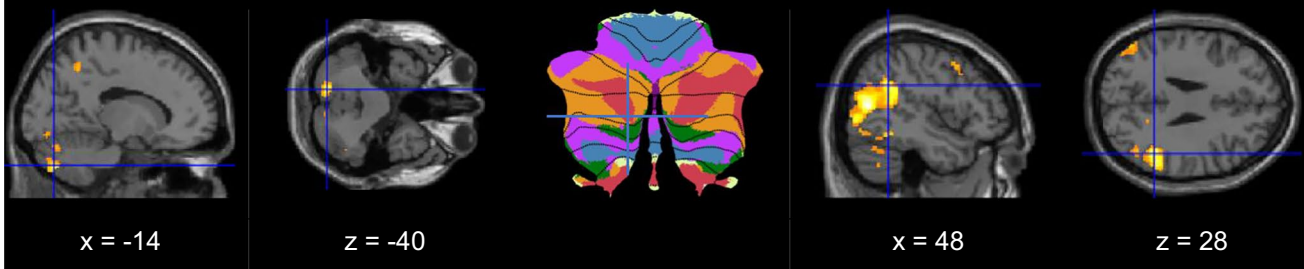
x = -14   z = -40   x = 48   z = 28

◄**Fig. 7** Spreading interaction showing higher cerebellar and TPJ activation at the explicit Belief SRT task at *general learning* (**a**), at *maintenance of learning* (**b**), and at *detecting violations* (**c**) compared with none at the explicit Control SRT task, displayed at an uncorrected threshold of $p < 0.005$, with color bars denoting SwE z values. Sagittal and transverse views of brain activations at the peak coordinates of the cerebellar Crus II [**Left**] and TPJ [**Right**]. The peak cerebellar activation is also indicated with a blue crosshair on flatmaps (Diedrichsen & Zotow, 2015) displaying functional networks by Buckner et al. (2011). Random block = Conjunction of Total Random and Random Orientation Blocks. Consistent with our hypotheses, compared to the explicit Control SRT task, the cerebellar Crus and TPJ show stronger activations in the learning effects of explicit Belief SRT task. Note that not all visible clusters are significant after FDR correction

cerebellar Crus II activation (Fig. 7b left), and the ROI analysis also revealed bilateral TPJ activation (Fig. 7b right). The whole-brain analysis further revealed activations in the calcarine gyrus, middle temporal gyrus, fusiform gyrus, precuneus, midcingulate cortex, paracentral lobule, postcentral gyrus, medial temporal pole, and superior medial gyrus.

Third, for the spreading interaction of *detecting violations,* the ROI and whole-brain analyses showed left cerebellar Crus II activation (Fig. 7c left), and the ROI analysis also revealed bilateral TPJ activation (Fig. 7c right). The whole-brain analysis further revealed activations in the middle temporal gyrus, precuneus, postcentral gyrus, precentral gyrus, middle frontal gyrus, and superior frontal gyrus.

To test our hypothesis more exhaustively, we also ran reverse *spreading interactions* specifying higher activations in the explicit Control SRT task than the explicit Belief SRT task. The ROI analysis showed an unexpected activation in the right Crus II during *maintenance of learning.* None of the other *a priori* ROIs were activated, consistent with our hypothesis. The whole-brain analysis did not reveal further activations of interest (Table 1).

Neuroimaging results: Comparing Explicit and Implicit Belief SRT tasks (cf. Hypothesis 2)

We next tested potential differences between results of the present explicit Belief SRT task and the previously published implicit Belief SRT task (Ma, Pu, et al., 2021b). To recall, these two tasks were identical except for the instruction mentioning (versus not mentioning, respectively) that there was a sequence.

Before moving to the fMRI results, we compared the behavioral results of the explicit and implicit Belief SRT tasks and found no significant differences with regards to error rates and RTs (see supplementary Fig. S3). For the sequence reproduction after scanning, participants who received explicit instructions about the presence of a sequence showed better performance than the participants who received the implicit instruction (Mean length: 4.45 for

explicit and 1.83 for implicit, $t = 3.61$, $p = 0.001$), indicating that participants did learn more about the sequence when they received explicit instructions about its presence.

Simple contrasts between explicit versus implicit Belief SRT tasks

We first conducted simple contrasts for all Block Types (Standard and Random) at all Phases (Training and Test) between explicit and implicit Belief SRT tasks. However, no significant activations were found.

Spreading interactions across explicit versus implicit Belief SRT tasks

We again applied a series of asymmetric or *spreading interactions,* now assuming higher activation in the implicit than in the explicit Belief SRT task for the three learning effects (Table 2 left panel). First, for the spreading interaction of *general learning*, the ROI analysis showed activation of the cerebellar Crus I (Fig. 8a), while the whole-brain analysis did not reveal significant activation. Second, for the spreading interaction of *maintenance of learning* (Fig. 8b), the ROI and whole-brain analyses showed activation of bilateral cerebellar Crus I/II, and the whole-brain analysis further revealed additional cerebellar activation in lobule IX, and cortical activations in the middle/superior temporal gyrus, caudate nucleus, and putamen. Third, for the spreading interaction of *detecting violations,* none of the ROI and whole-brain analyses showed significant activations.

To test our hypothesis more exhaustively, we also ran reverse *spreading interactions* specifying higher activation in the three learning effects in the explicit rather than the implicit Belief SRT task (Table 2 right panel). For the spreading interaction of *general learning*, no significant activations were found. However, contrary to the hypothesis, for the spreading interaction of *maintenance of learning* (Fig. 8c), the ROI analysis showed activation of the right cerebellar Crus II. Thus, both in the implicit and explicit SRT Belief task, cerebellar activation is stronger during Test than during Training, but in slightly different nonoverlapping areas involving greater activation in the bilateral Crus I/II and lobule IX in the implicit task, and only in the right Crus II in the explicit task. For the spreading interaction of *detecting violations*, the ROI analysis showed significant TPJ activations (Fig. 8d).

Discussion

This study investigated, for the first time, how the cerebellar Crus and the cortical TPJ contribute to explicit learning of a belief sequence in a serial response time task (i.e.,
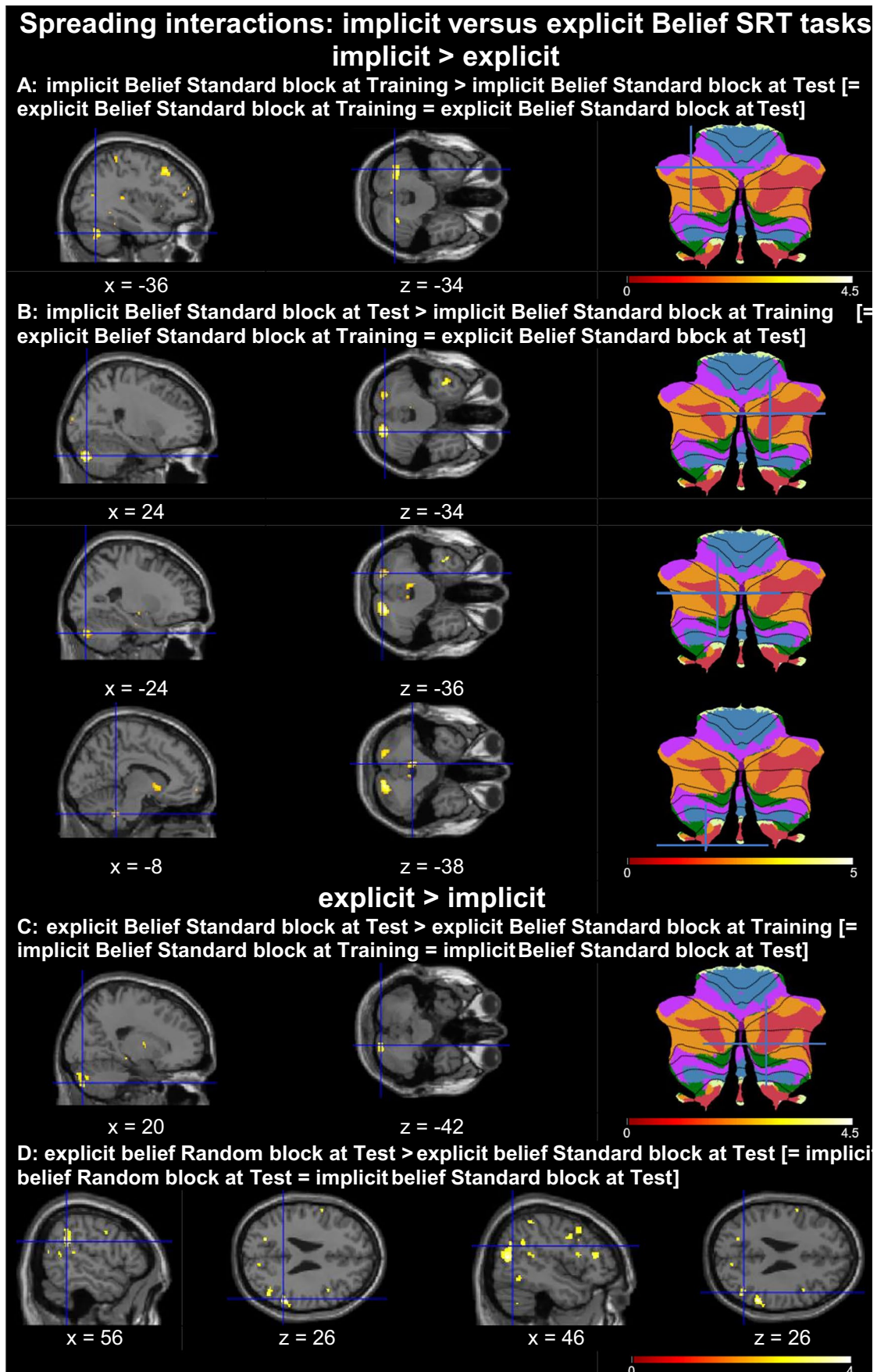
Spreading interactions: implicit versus explicit Belief SRT tasks

implicit > explicit

A: implicit Belief Standard block at Training > implicit Belief Standard block at Test [= explicit Belief Standard block at Training = explicit Belief Standard block at Test]

x = -36   z = -34   0 ————————— 4.5

B: implicit Belief Standard block at Test > implicit Belief Standard block at Training   [= explicit Belief Standard block at Training = explicit Belief Standard block at Test]

x = 24   z = -34

x = -24   z = -36

x = -8   z = -38   0 ————————— 5

explicit > implicit

C: explicit Belief Standard block at Test > explicit Belief Standard block at Training [= implicit Belief Standard block at Training = implicit Belief Standard block at Test]

x = 20   z = -42   0 ————————— 4.5

D: explicit belief Random block at Test > explicit belief Standard block at Test [= implicit belief Random block at Test = implicit belief Standard block at Test]

x = 56   z = 26   x = 46   z = 26   0 —————— 4

◄**Fig. 8** Spreading interaction showing higher cerebellar activation at the implicit Belief SRT task at *general learning* (**a**), at *maintenance of learning* (**b**) compared with none at the explicit Belief SRT task. Spreading interaction showing higher cerebellar and TPJ activation at the explicit Belief SRT task at *maintenance of learning* (**c**), and at *detecting violations* (**d**) compared with none at the implicit Belief SRT task. All figures were displayed at an uncorrected threshold of *p* < 0.005, with color bars denoting SwE z values. Sagittal and transverse views of the cerebellar and TPJ activations at the peak coordinates, indicated by blue crosshairs. The peak cerebellar activation is also indicated with a blue crosshair on flatmaps (Diedrichsen & Zotow, 2015) displaying functional networks by Buckner et al. (2011). Random block = Conjunction of Total Random and Random Orientation Blocks. Note that not all visible clusters are significant after FDR correction

explicit Belief SRT task). In this task, participants were explicitly informed that there was a sequence in the true and false belief trials presented to them. Two main results are noteworthy.

First, participants' behavioral performance and brain activation were compared against an explicit Control SRT task, which consisted of the same task and sequence structure, but with non-social elements. Overall, as hypothesized, the bilateral cerebellar Crus II and the cortical TPJ were more strongly activated in the explicit Belief SRT than in the explicit Control SRT task. In particular, the posterior cerebellar Crus II was activated during maintenance of the learned belief sequence and detection of violations in the Test phase, and the cortical TPJ was activated during all phases of the explicit Belief SRT task, including general sequence learning in the initial Training phase, maintaining the learned sequence and detecting violations in the Test phase (Table 1; see summary Table 3). Moreover, we found that faster response times were correlated with higher activation of the cerebellar Crus I & II and lower activation of the TPJ. In contrast, a small area of Crus II was engaged more during sequence maintenance in the explicit Control than in the Belief SRT task. However, there was no relationship between this Crus II activation and response times in the explicit Control SRT task.

Second, a comparison was made between the current explicit Belief SRT task and an implicit version of the same task used in previous research (i.e., without informing about the existence of a standard sequence, or "implicit" Belief SRT task; Ma, Pu, et al., 2021b). The results showed that the posterior cerebellum was engaged more during implicit learning. Specifically, the cerebellar Crus I was engaged more during general sequence learning in the initial Training phase, and the bilateral cerebellar Crus I & II and lobule IX were engaged more when maintaining this sequence during the Test phase. In contrast, during explicit belief learning, a distinct, nonoverlapping area of the right Crus II was engaged during maintenance; the cortical TPJ was engaged more in detecting violations (Table 2; see summary Table 3).

Explicit Belief SRT task

With respect to the present explicit Belief SRT task, as hypothesized, we found that the bilateral cerebellar Crus I & II were highly activated while maintaining the learned Standard belief sequence during the Test phase (when interrupted by random sequences) compared with the initial Training phase. Moreover, to explore the relationship between the neural time course of the posterior cerebellar Crus II and reaction times throughout the explicit Belief SRT task, a Spearman correlation was calculated. The results showed a relationship between faster response times and increased cerebellar Crus activation across the whole Belief SRT task. This suggests that the repeated exposure to the sequence in the Belief SRT task increased activation in Crus I & II, reflecting firmer encoding and automatization of the sequence, which in turn improved behavioral performance (i.e., faster RTs). This is consistent with the idea, put forward by Bernard and Seidler (2013), that stronger cerebellar activation during sequence learning is indicative of a newly formed internal model of the embedded sequence. The increased activation of cerebellar Crus thus suggested successful learning and increasingly automatic and faster application of the repeated belief sequence. Taken together, our results favor the hypothesis that the posterior cerebellum supports social functions by predicting and automatizing one´s own and others´ social actions (Van Overwalle, Manto, et al., 2019b).

Although activation of cerebellar Crus II also was found in the Control SRT task, the clusters in the Belief SRT task were larger and more consistently activated. Moreover, as noted earlier, a correlation with response times was only found in the Belief SRT task, which indicates that sequences were only firmly stored in Crus II under a social task context. This suggests that Crus II may be sensitive to all sorts of sequences, but preferentially for social sequences. This interpretation is further supported given that Crus II is located in the default/mentalizing network (Buckner et al., 2011; Van Overwalle, Ma, & Heleven, 2020a). Together, these observations support our hypothesis that the posterior cerebellar Crus is preferentially involved in social contexts (Guell et al., 2018; Hoche et al., 2016; Van Overwalle, Ma, & Heleven, 2020a).

Alternatively, one could argue that higher posterior cerebellar activation in the explicit Belief SRT task compared with the explicit Control SRT task, was not due to the social context, but rather related to better behavioral performance in the explicit Belief SRT task. Indeed, participants showed faster response times and more accurate post-scan sequence reproduction in the explicit Belief SRT task than the Control SRT task. Although this argument is reasonable, it seems to ignore the possibility that better behavioral performance in the explicit Belief SRT task may, in fact, be due also to the

Table 3 Summary of ROI results

Contrasts	Within task		Between tasks (Spreading interactions)			
	Explicit (Table S1)		Explicit (Table 1)		Implicit vs. explicit (Table 2)	
Learning effects	Belief	Control	Belief > Control	Control > Belief	Implicit > Explicit	Explicit > Implicit
General learning	---	---	---	---	Crus I	---
	R TPJ	---	R TPJ	---	---	---
Maintenance of learning	R/L Crus II	R/L Crus II	R/L Crus II	R Crus II	R/L Crus II	R Crus II
	---	---	R/L TPJ	---	---	---
Detecting violations	---	---	L Crus II	---	---	---
	R TPJ	---	R/L TPJ	---	---	R TPJ
			Between Tasks (Simple contrasts)			
			Explicit (Table S2)		Implicit vs. Explicit	
Simple contrasts			Belief > Control	Control > Belief	Implicit > Explicit	Explicit > Implicit
Standard block at training			---	---	---	---
			R TPJ	---	---	---
Standard block at test			L Crus II	---	---	---
			R/L TPJ	---	---	---
Total random at test			L Crus II	---	---	---
			R/L TPJ	---	---	---
Random orientation at test			L Crus II	---	---	---
			R TPJ	---	---	---

Note: *General learning* = Standard block at Training > Standard block at Test, *Maintenance of learning* = Standard block at Test > Standard block at Training, *Detecting violations* = Total Random block at Test > Standard block at Test & Random Orientation block at Test > Standard block at Test. L = Left, R = Right. "---" indicates no activation in the ROI

social context, as people are more familiar with social compared to cognitive stimuli (Smurfs vs. shapes), and people have more extensive practice on belief inference during daily life as opposed to artificial rules based on colored shapes (Callejas et al., 2011; Cohen & German, 2010). We saw that participants improved performance (i.e., faster responses) during repeated presentations of the standard sequence in both the explicit Belief and Control SRT tasks. If the behavioral performance argument would be correct, faster response times should correlate with changes of posterior cerebellar activation in both tasks. However, this was not the case. Correlations were observed only for the explicit Belief SRT task and not for the explicit Control SRT task. This indicates that the posterior cerebellum contributed more to responding to a belief sequence and less to other sorts of sequences, providing support for our social interpretation. Furthermore, if the behavioral performance argument is correct, participants should show stronger cerebellar activation in the explicit SRT tasks compared with the implicit SRT tasks as they should be able to recollect the explicit sequence better. This was not the case either (Table 2, see detailed discussion in the next section). Taking all these observations into account, our study demonstrated that the posterior cerebellum is preferentially involved in social processes.

Parallel to the posterior cerebellum, and consistent with our hypothesis, the cortical TPJ was consistently and preferentially activated more in the explicit Belief SRT task than the explicit Control SRT task. This is in line with previous research showing TPJ activation when implicitly learning a sequence of true and false beliefs (Ma, Pu, et al., 2021b), and with established knowledge that the TPJ serves the role of switching to another's perspective and inferring their transient mental states (Schurz et al., 2013; Van Overwalle, 2009).

Interestingly, in the explicit Belief SRT task, the TPJ was activated when learning the belief sequence at the initial Training phase and when detecting violations in the learned sequence, but not while maintaining the sequence at the Test phase (Table S2). Correlation analysis further revealed a pattern of faster response times and decreased cortical TPJ activation during standard belief sequences, and an opposite pattern of longer response times and increased TPJ activation during violations of the belief sequence. This suggests that the repeated exposure to the embedded sequence resulted in applying the newly formed internal model in Crus II (as mentioned earlier), and so aided in anticipating the next trial, which led to decreased reaction times for perspective switching and decreased TPJ activation for making true or false belief inferences (Koster-Hale & Saxe, 2013). When there

was an unexpected random sequence trial and anticipations and observed beliefs did not match, the TPJ was activated to reorient attention to unexpected beliefs and infer protagonists' new beliefs (Geng & Vossel, 2013; Koster-Hale & Saxe, 2013). Overall, this suggested that the TPJ sends social information to help in building and adjusting an internal model mainly at the initial training and when there are violations.

Also, our results of the explicit Belief SRT task showed that the right TPJ was consistently activated throughout all comparisons (except *maintenance of learning*), while the left TPJ was activated only in direct comparisons between explicit Belief and Control SRT tasks (summary in Table 3). These results are in line with meta-analyses which indicated that both left and right TPJ are activated during mentalizing tasks, although the right TPJ might be more prominently activated (Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle & Baetens, 2009). However, hemispheric differences in TPJ activation during belief inferences are still poorly understood so that we cannot draw any firm conclusions about these results.

The significant correlations between cerebellar Crus II, cortical TPJ, and reaction times support previous anatomical evidence of closed cerebro-cerebellar loops: i.e., bidirectional connections between the cerebellum and identical areas to and from the cerebral cortex (Kelly & Strick, 2003). Research in the social domain demonstrated that there is functional closed-loop connectivity between the posterior cerebellar Crus and cortical TPJ during sequencing stories involving beliefs (Van Overwalle, Van de Steen, et al., 2020b). The result shows that input from the cerebrum (e.g., the TPJ) projects to areas of the cerebellum (e.g., Crus I & II), which in turn send error information back to the same TPJ areas (Van Overwalle, Van de Steen, et al., 2020b). Incidentally, very similar posterior Crus-TPJ closed-loop connectivity was found in other mentalizing tasks, such as trait inferences of persons and stereotypes inferences of groups (Van Overwalle, Van de Steen, & Mariën, 2019c). Future research could investigate the functional connectivity between the posterior cerebellar Crus and cortical TPJ in the present context of social sequence learning.

Additional cortical activations were observed during belief sequence learning (Tables 1 & S2). The dorsal premotor cortex possibly reflects selecting and updating appropriate responses according to the observed input and inferred true or false belief (Hardwick et al., 2013). Our current study also yielded activation in the caudate nucleus, pallidum, and putamen, which have been implicated in both implicit and explicit learning and fine-tuning of sequential information (Baetens et al., 2020; Janacsek et al., 2020). Previously, these areas have been linked to motor sequencing (Doyon et al., 2009) and perceptual sequencing (Turk-Browne et al., 2009). Our results suggest that these areas also are involved

in belief sequencing. This supports a domain-general role of these areas in sequence learning. In addition, the Belief SRT task activated several critical regions in the mentalizing network, including the TPJ, which is known for supporting belief reasoning (Van Overwalle & Baetens, 2009), and the precuneus, which is linked to constructing mental images related to different perspectives in social settings (Frith & Frith, 2006; Schurz et al., 2013). This latter result provides further support for the hypothesis by Van Overwalle and Baetens (2009) that the precuneus contributes to the mentalizing process by retrieving learned situations and matching them with the current situations in order to prepare appropriate perspective taking. Future research could investigate the connectivity between the posterior cerebellum and these cerebral areas related to sequence learning and metalizing.

Comparing the Explicit and Implicit Belief SRT tasks

To explore potential differences and similarities between explicit and implicit sequence learning in a social context, we further compared the current explicit Belief SRT task to the implicit Belief SRT task applied in previous research (Ma, Pu, et al., 2021b). Spreading interactions showed that the posterior cerebellum was more strongly activated in implicit than in explicit learning. In particular, the cerebellar Crus I was engaged more during initial learning, and the bilateral cerebellar Crus I & II and lobule IX were engaged more when maintaining this sequence during testing. Although the reverse interaction identified an area of the right cerebellar Crus II that was activated more during explicit than implicit learning, overall, the active cerebellar clusters were much larger during implicit learning (cf. whole brain analysis; Table 2). One possible explanation, consistent with our hypothesis, is that the posterior cerebellum is activated while storing belief sequences no matter whether participants learned the sequence implicitly or explicitly, more so during implicit learning (Morgan et al., 2020). Another possible explanation is that there might be a functional differentiation within the cerebellar Crus II related to implicit versus explicit social sequence learning. Together, these findings suggest posterior cerebellar involvement in social contexts in both implicit and explicit sequence processing, and future research is needed to investigate to what extent areas within Crus II are distinctly involved in implicit or explicit learning.

One interesting finding is the cerebellar lobule IX was activated more while maintaining the belief sequence in the implicit Belief SRT task. As shown in Figure 8b, this activation is also located in the default/mentalizing network identified by Buckner et al. (2011). Haihambo et al. (2021) also found that the cerebellar lobule IX was engaged in predictions of temporal sequences of others' social actions. These

findings support the idea, put forward by Guell et al. (2018), that the cerebellar lobule IX/X constitutes a third representation of social cognitive processing, alongside Crus I and II as primary and secondary representations. Because there is little evidence in the current literature on the social function of lobule IX, its role in social sequencing is unclear and requires future investigation.

Although participants received identical instructions about how to infer protagonists' beliefs (i.e., oriented towards vs. away from the screen implying true vs. false beliefs), which requires a similar level of mentalizing across implicit and explicit Belief SRT tasks, the results showed stronger TPJ activation when detecting violations in the explicit Belief SRT task. This may be due to the explicit instruction activating extracerebellar pathways to help the sequence learning process (Morgan et al., 2020; Taylor et al., 2010), resulting in higher activation of the TPJ. Another explanation is that participants in the explicit task were somewhat more aware of the standard sequence and therefore exerted more mental effort and TPJ activation to adjust to violations.

Note that in the current study, participants did not learn the sequence before entering the scanner. So, while they were explicitly informed about the existence of the standard sequence, it is likely that they would not be able to reproduce the entire sequence. This also may be an additional reason why implicit and explicit Belief SRT tasks do not differ much. We chose to provide this limited explicit instruction, because telling participants the exact sequence would reflect sequence retrieval rather than learning (Deroost & Coomans, 2018). It will, however, be of interest for future research to investigate brain activations at different levels of learning (e.g., with full, partial or no knowledge of the sequence).

Implications

Our results revealed that the posterior cerebellum was preferentially engaged during sequencing of social information, which provides potential insights for clinical research on cerebellar dysfunctions and stimulation. Moreover, we found that cerebellar Crus I & II were less strongly activated in explicit than in implicit learning, in line with the cerebellar role in automatization. However, because we only recruited healthy participants, it is possible that there are more differences between implicit and explicit belief sequence learning for patients with any kind of cerebellar damage (e.g., cerebellar patients, individuals with autism, etc.; D'Mello & Stoodley, 2015). This should be investigated in future research.

Our results showed that the posterior cerebellum is less involved in explicit than in implicit sequence learning. As noted earlier, this weaker activation may be due to the fact that explicit strategies may activate extracortical

areas to help the learning process. However, there is still reason to suspect that explicit strategies cannot sufficiently compensate for clinical impairments in sequence learning in social life, especially when the complexity of social input increases (Frith, 2012; Siciliano & Clausi, 2020). For example, as mentioned earlier, cerebellar patients are capable of identifying the correct chronological order of routine social actions, but not of social actions that require inferring peoples' mental states (Van Overwalle, De Coninck, et al., 2019a), and they also fail in advanced mentalizing tasks (Clausi et al., 2019). Indeed, successful social interactions rely on detecting patterns of multiple types of social information, such as gestures, facial expressions, and eye contact (Lieberman, 2000), which is more complex than what we manipulated in our research. For this reason, future SRT tasks could embed more complex social cues and information to investigate the potential limiting conditions of efficient implicit and explicit social sequence learning.

The present results support the idea that the posterior cerebellum might be a novel target for noninvasive neurostimulation in the social domain. Previous stimulation studies typically targeted the whole cerebellum, and found that anodal cerebellar transcranial direct current stimulation (tDCS) improved learning of sequential stimuli (e.g., dots on the screen) by healthy participants (Ballard et al., 2019; Ferrucci et al., 2013). The current findings suggest that the cerebellar Crus I and II play a domain-specific role in learning and automatizing social sequences and, therefore, can be a potential target for noninvasive neurostimulation and neuro-guided therapy/training for patients with cerebellar lesions who have impairments in social cognition (Clausi et al., 2019; Van Overwalle, De Coninck, et al., 2019a). This approach might be extended and applied to other clinical dysfunctions that are largely caused by cerebellar impairments, such as autism (D'Mello & Stoodley, 2015). Importantly, a unique contribution of the current Belief SRT task in future noninvasive stimulation studies is that it may elucidate whether brain stimulation can improve both implicit and explicit learning of social sequences.

Conclusions

The current study shows the critical role of the posterior cerebellar Crus I & II and the cortical TPJ in explicit belief sequence learning. Notably, we found that the posterior cerebellar Crus I & II was activated in maintaining an embedded belief sequence and that the TPJ was activated to support in capturing and updating true versus false belief information. Additionally, for the first time, this study revealed insightful correlations of posterior cerebellar Crus and TPJ activation with response times

during explicit learning of belief sequences, indicating that the cerebellar Crus becomes more active and the TPJ become less active when sequence learning progresses, whereas Crus and TPJ both become more active when violations occur. The current study also revealed that, compared with implicit learning, explicit belief sequence learning engaged the cerebellum less but the cortical TPJ more. Future studies could investigate potential differences between implicit and explicit social sequence learning on patients with cerebellar damage or social impairments and could explore the effect of posterior cerebellar stimulation (e.g., using tDCS) on implicit and explicit social sequence learning.

References

Baetens, K., Firouzi, M., Van Overwalle, F., & Deroost, N. (2020). Involvement of the cerebellum in the serial reaction time task (SRT) (Response to Janacsek et al.). *NeuroImage*, *220*, 117114. https://doi.org/10.1016/j.neuroimage.2020.117114

Ballard, H. K., Goen, J. R. M., Maldonado, T., & Bernard, J. A. (2019). Effects of cerebellar transcranial direct current stimulation on the cognitive stage of sequence learning. *Journal of Neurophysiology*, *122*(2), 490–499. https://doi.org/10.1152/jn.00036.2019

Bernard, J. A., & Seidler, R. D. (2013). Cerebellar contributions to visuomotor adaptation and motor sequence learning: An ALE meta-analysis. *Frontiers in Human Neuroscience*, *7*, 27. https://doi.org/10.3389/fnhum.2013.00027

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(5), 2322–2345. https://doi.org/10.1152/jn.00339.2011

Callejas, A., Shulman, G. L., & Corbetta, M. (2011). False Belief vs. False Photographs: A Test of Theory of Mind or Working Memory? *Frontiers in Psychology*, *2*, 316. https://doi.org/10.3389/fpsyg.2011.00316

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243–1249. https://doi.org/10.1093/cercor/bhi007

Clausi, S., Olivito, G., Lupo, M., Siciliano, L., Bozzali, M., & Leggio, M. (2019). The cerebellar predictions for social interactions: Theory of mind abilities in patients with degenerative cerebellar atrophy. *Frontiers in Cellular Neuroscience*, *12*, 510. https://doi.org/10.3389/fncel.2018.00510

Cohen, A. S., & German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for ' theory of mind'. *Cognition*, *115*(3), 417–425. https://doi.org/10.1016/j.cognition.2010.03.001

Coomans, D., Deroost, N., Zeischka, P., & Soetens, E. (2011). On the automaticity of pure perceptual sequence learning. *Consciousness and Cognition*, *20*(4), 1460–1472. https://doi.org/10.1016/j.concog.2011.06.009

Cusack, R., & Papadakis, N. (2002). New robust 3-D phase unwrapping algorithms: Application to magnetic field mapping and undistorting echoplanar images. *NeuroImage*, *16*(3), 754–764. https://doi.org/10.1006/nimg.2002.1092

D'Mello, A. M., & Stoodley, C. J. (2015). Cerebro-cerebellar circuits in autism spectrum disorder. *Frontiers in Neuroscience*, *9*, 408. https://doi.org/10.3389/fnins.2015.00408

Debas, K., Carrier, J., Orban, P., Barakat, M., Lungu, O., Vandewalle, G., … Doyon, J. (2010). Brain plasticity related to the consolidation of motor sequence learning and motor adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(41), 17839–17844. https://doi.org/10.1073/pnas.1013176107

Deroost, N., & Coomans, D. (2018). Is sequence awareness mandatory for perceptual sequence learning: An assessment using a pure perceptual sequence learning design. *Acta Psychologica*, *183*, 58–65. https://doi.org/10.1016/j.actpsy.2018.01.002

Diedrichsen, J., & Zotow, E. (2015). Surface-based display of volume-averaged cerebellar imaging data. *PLoS ONE*, *10*(7), 1–18. https://doi.org/10.1371/journal.pone.0133402

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., … Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research*, *199*(1), 61–75. https://doi.org/10.1016/j.bbr.2008.11.012

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferrucci, R., Brunoni, A. R., Parazzini, M., Vergari, M., Rossi, E., Fumagalli, M., … Priori, A. (2013). Modulating Human Procedural Learning by Cerebellar Transcranial Direct Current Stimulation. *The Cerebellum*, *12*(4), 485–492. https://doi.org/10.1007/s12311-012-0436-9

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Gao, Y., Rogers, J. C., Pauli, R., Clanton, R., Baker, R., Birch, P., … De Brito, S. A. (2019). Neural correlates of theory of mind in typically-developing youth: Influence of sex, age and callous-unemotional traits. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-52261-y

Geiger, A., Cleeremans, A., Bente, G., & Vogeley, K. (2018). Social Cues Alter Implicit Motor Learning in a Serial Reaction Time Task. *Frontiers in Human Neuroscience*, *12*, 1–12. https://doi.org/10.3389/fnhum.2018.00197

Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: Contextual updating? *Neuroscience and Biobehavioral Reviews*, *37*(10), 1–13. https://doi.org/10.1016/j.neubiorev.2013.08.010

Guell, X., Gabrieli, J. D. E., & Schmahmann, J. D. (2018). Triple representation of language, working memory, social and emotion processing in the cerebellum: convergent evidence from task and seed-based resting-state fMRI analyses in a single large cohort.

NeuroImage, *172*, 437–449. https://doi.org/10.1016/j.neuroimage.2018.01.082

Guillaume, B., Hua, X., Thompson, P. M., Waldorp, L., & Nichols, T. E. (2014). Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. *NeuroImage*, *94*, 287–302. https://doi.org/10.1016/j.neuroimage.2014.03.029

Haihambo, N., Ma, Q., Baeken, C., Deroost, N., Baetens, K., Heleven, E., & Van Overwalle, F. (2021). Social Thinking is for Doing: The Posterior Cerebellum Supports Prediction of Social Actions Based on Personality Traits. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsab087

Hardwick, R. M., Rottschy, C., Miall, R. C., & Eickhoff, S. B. (2013). A quantitative meta-analysis and review of motor learning in the human brain. *NeuroImage*, *67*, 283–297. https://doi.org/10.1016/j.neuroimage.2012.11.020

Heleven, E., van Dun, K., & Van Overwalle, F. (2019). The posterior Cerebellum is involved in constructing Social Action Sequences: An fMRI Study. *Scientific Reports*, *9*(1), 11110. https://doi.org/10.1038/s41598-019-46962-7

Hoche, F., Guell, X., Sherman, J. C., Vangel, M. G., & Schmahmann, J. D. (2016). Cerebellar Contribution to Social Cognition. *The Cerebellum*, *15*(6), 732–743. https://doi.org/10.1007/s12311-015-0746-9

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, *9*(4), 304–313. https://doi.org/10.1038/nrn2332

Janacsek, K., Shattuck, K. F., Tagarelli, K. M., Lum, J. A. G., Turkeltaub, P. E., & Ullman, M. T. (2020). Sequence learning in the human brain: A functional neuroanatomical meta-analysis of serial reaction time studies. *NeuroImage*, *207*, 116387. https://doi.org/10.1016/j.neuroimage.2019.116387

Kelly, R. M., & Strick, P. L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *Journal of Neuroscience*, *23*(23), 8432–8444. https://doi.org/10.1523/jneurosci.23-23-08432.2003

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, *79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

Leggio, M., & Molinari, M. (2015). Cerebellar Sequencing: a Trick for Predicting the Future. *The Cerebellum*, *14*(1), 35–38. https://doi.org/10.1007/s12311-014-0616-x

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin*, *126*(1), 109–136. https://doi.org/10.1037/0033-2909.126.1.109

Ma, Q., Heleven, E., Funghi, G., Pu, M., Baetens, K., Deroost, N., & Van Overwalle, F. (2021a). Implicit Learning of True and False Belief Sequences. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.643594

Ma, Q., Pu, M., Heleven, E., Haihambo, N. P., Baetens, K., Baeken, C., … Van Overwalle, F. (2021b). The posterior cerebellum supports implicit learning of social belief sequences. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-021-00910-z

Mastroianni, A. M., Gilbert, D. T., Cooney, G., & Wilson, T. D. (2021). Do conversations end when people want them to? *Proceedings of the National Academy of Sciences*, *118*(10), e2011809118. https://doi.org/10.1073/pnas.2011809118

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Morgan, O. P., Slapik, M. B., Iannuzzelli, K. G., LaConte, S. M., Lisinski, J. M., Nopoulos, P. C., … Marvel, C. L. (2020). The Cerebellum and Implicit Sequencing: Evidence from Cerebellar Ataxia. *Cerebellum*. https://doi.org/10.1007/s12311-020-01206-7

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Pu, M., Heleven, E., Delplanque, J., Gibert, N., Ma, Q., Funghi, G., & Van Overwalle, F. (2020). The posterior cerebellum supports the explicit sequence learning linked to trait attribution. *Cognitive, Affective and Behavioral Neuroscience*, *20*(4), 798–815. https://doi.org/10.3758/s13415-020-00803-7

Pu, M., Ma, Q., Heleven, E., Haihambo, N. P., & Van Overwalle, F. (2021). The posterior cerebellum and inconsistent trait implications when learning the sequence of actions. *Social Cognitive and Affective Neuroscience*, 1–11. https://doi.org/10.1093/scan/nsab037

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, *7*, 712. https://doi.org/10.3389/fnhum.2013.00712

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Siciliano, L., & Clausi, S. (2020). Implicit vs. Explicit Emotion Processing in Autism Spectrum Disorders: An Opinion on the Role of the Cerebellum. *Frontiers in Psychology*, *11*, 1–8. https://doi.org/10.3389/fpsyg.2020.00096

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage*, *35*(3), 1378–1384. https://doi.org/10.1016/j.neuroimage.2007.01.042

Taylor, J. A., Klemfuss, N. M., & Ivry, R. B. (2010). An explicit strategy prevails when the cerebellum fails to compute movement errors. *Cerebellum*, *9*(4), 580–586. https://doi.org/10.1007/s12311-010-0201-x

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/jocn.2009.21131

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584. https://doi.org/10.1016/j.neuroimage.2009.06.009

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *NeuroImage*, *86*, 554–572. https://doi.org/10.1016/j.neuroimage.2013.09.033

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2015). Cerebellar areas dedicated to social cognition? A comparison of meta-analytic and connectivity results. *Social Neuroscience*, *10*(4), 337–344. https://doi.org/10.1080/17470919.2015.1005666

Van Overwalle, F., De Coninck, S., Heleven, E., Perrotta, G., Taib, N. O. Ben, Manto, M., & Mariën, P. (2019a). The role of the cerebellum in reconstructing social action sequences: A pilot

study. *Social Cognitive and Affective Neuroscience*, *14*(5), 549–558. https://doi.org/10.1093/scan/nsz032

Van Overwalle, F., Manto, M., Leggio, M., & Delgado-García, J. M. (2019b). The sequencing process generated by the cerebellum crucially contributes to social interactions. *Medical Hypotheses*, *128*, 33–42. https://doi.org/10.1016/j.mehy.2019.05.014

Van Overwalle, F., Van de Steen, F., & Mariën, P. (2019c). Dynamic causal modeling of the effective connectivity between the cerebrum and cerebellum in social mentalizing across five studies. *Cognitive, Affective and Behavioral Neuroscience*, *19*(1), 211–223. https://doi.org/10.3758/s13415-018-00659-y

Van Overwalle, F., Ma, Q., & Heleven, E. (2020a). The Posterior Crus II Cerebellum is specialized for Social Mentalizing and Emotional self-Experiences: A Meta-analysis. *Social Cognitive and Affective Neuroscience*. https://doi.org/10.1093/scan/nsaa124

Van Overwalle, F., Van de Steen, F., van Dun, K., & Heleven, E. (2020b). Connectivity between the cerebrum and cerebellum during social and non-social sequencing using dynamic causal modelling. *NeuroImage*, *206*, 116326. https://doi.org/10.1016/j.neuroimage.2019.116326