

# Structural Maturation and Brain Activity Predict Future Working Memory Capacity during Childhood Development

Henrik Ullman, Rita Almeida, and Torkel Klingberg

Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden

Human working memory capacity develops during childhood and is a strong predictor of future academic performance, in particular, achievements in mathematics and reading. Predicting working memory development is important for the early identification of children at risk for poor cognitive and academic development. Here we show that structural and functional magnetic resonance imaging data explain variance in children's working memory capacity 2 years later, which was unique variance in addition to that predicted using cognitive tests. While current working memory capacity correlated with frontoparietal cortical activity, the future capacity could be inferred from structure and activity in basal ganglia and thalamus. This gives a novel insight into the neural mechanisms of childhood development and supports the idea that neuroimaging can have a unique role in predicting children's cognitive development.

## Introduction

Development of cognitive functions during childhood and adolescence is an important foundation for academic achievement and social functioning. In particular, working memory (WM) capacity, the ability to keep information online for a short period of time for cognitive processing, is one of the strongest predictors of future achievements in mathematics and reading (Gathercole et al., 2004; Dumontheil and Klingberg, 2012). Neuroimaging could possibly have a role in identifying children at risk for poor cognitive development, but it is unclear whether neuroimaging contributes above and beyond that which can be predicted from cognitive tests (Gabrieli, 2009).

Magnetic resonance imaging (MRI) allows noninvasive measurement of both structural and functional brain development in children. Several studies have used MRI to show developmental changes in working memory (WM)-related brain activation (Klingberg et al., 2002; Kwon et al., 2002; Crone et al., 2006; Scherf et al., 2006) but not to predict future WM capacity. Machine learning algorithms, based on multivariate data, are promising for this goal (Haxby et al., 2001). Adaptations of machine learning algorithms for regression problems have resulted in methods suited for making quantitative predictions from noisy multivariate data. One such approach is Support Vector Regression (SVR) (Scholkopf et al., 1999), which has been used to esti-

mate the age of children from patterns of brain activity (Dosenbach et al., 2010). Here we used the same approach and a longitudinal design to investigate the potential for neuroimaging to predict future cognitive capacity, in particular, performance on a visuospatial WM task. Next, we tested whether neuroimaging can contribute to the prediction beyond what can be inferred from cognitive tests.

## Materials and Methods

**Participants.** Subjects in the current study were randomly selected from a larger pool of healthy volunteers ( $n = 323$ ) recruited using random sampling from the population registry in the town of Nynäshamn in Sweden (Söderqvist et al., 2010). Subjects were recruited in age groups from 6 to 20 years with a 2 year gap between the groups. The maximum age span at the time of testing within each group was 7 months. The total number of subjects in the sample used to train the SVR model was 62 (30 females). The sizes of the age groups were the following: aged 6,  $n = 11$ ; 8,  $n = 10$ ; 10,  $n = 9$ ; 12,  $n = 8$ ; 14,  $n = 9$ ; 16,  $n = 8$ ; 18,  $n = 4$ ; 20,  $n = 3$ . Exclusion criteria were a diagnosed neuropsychological disorder other than attention deficit hyperactive disorder (ADHD) and dyslexia. ADHD symptoms were expected at the same rate as the general population. One of the subjects had a diagnosis of dyslexia while none had a diagnosis of ADHD. The study was approved by the ethics committee at Karolinska University Hospital, Stockholm. Analyses of the relationship of WM and age were done using all subjects in the larger pool of data.

**Cognitive testing.** The subjects completed neuropsychological assessment including tests of reasoning and WM. This assessment was completed at two time points (t1 and t2) with 2 years between them. The current study used three tests of WM, as follows. (1) Visuospatial WM (VSWM) task, a dot matrix task from the Automated Working Memory Assessment (AWMA; Alloway, 2007). The participant watches a sequence of dots presented on a  $4 \times 4$  grid and is required to remember and recall the order and position of dots. The score used for the analysis is the sum of correct trials in this test. (2) Backwards digit task. The subject listens to a list of numbers and is required to recall the list in a backwards order. The difficulty level increases by adding one digit to the list after four correct trials, and terminating when three errors are committed in the same block of trials. (3) The 3-back task, where subjects are read a list of 20 words and after each word reports "yes" if it was the same as the word read three words earlier or "no" if it was different from the word

Received Feb. 24, 2013; revised Nov. 13, 2013; accepted Nov. 26, 2013.

Author contributions: H.U. and T.K. designed research; H.U. and R.A. performed research; H.U. analyzed data; H.U., R.A., and T.K. wrote the paper.

This work was supported by the Knut and Alice Wallenberg Foundation, The Swedish Research Council, and a Swedish Royal Bank Centennial Foundation Grant in the program, "Learning and Memory in Children and Young Adults" to T.K. The study was supported by the Karolinska Institutet Strategic Neuroscience Program to R.A. We thank Henrik Ehrsson, Hans Forssberg, Lars Nyberg and Megan Spencer-Smith for feedback regarding the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Henrik Ullman, Karolinska Institutet, Department of Neuroscience, Retzius Väg 8, 171 77 Stockholm, Sweden. E-mail: Henrik.Ullman@ki.se.

DOI:10.1523/JNEUROSCI.0842-13.2014

Copyright © 2014 the authors 0270-6474/14/341592-07\$15.00/0

read three words earlier. A summary score was calculated by subtracting the number of wrong “yes” responses from the number of correct responses. Finally, a composite score of working memory was constructed based on t1 data from the three WM tasks described. The composite measure was obtained using structural equation modeling and assuming that there is a working memory ability which is not directly measurable but is reflected in the scores obtained in individual tasks. The Raven’s Progressive Matrices test was used to measure reasoning ability (Raven et al., 1998). The 6-year-old age group performed subtests A–D, while the other groups performed subtests A–E. Raw scores from the WM tests were used in the statistical analysis, while the reasoning test scores were analyzed with item-response-theory (IRT). The IRT analysis resulted in a z-score that was comparable across the group. The test correlations between the two time points were the following: VSWM  $r = 0.75$ ,  $p < 0.001$ , Backwards digit task  $r = 0.74$ ,  $p < 0.001$ , 3-back task  $r = 0.19$ ,  $p = 0.19$ , Raven’s Progressive Matrices test  $r = 0.81$ ,  $p < 0.001$ .

**MRI measurements.** Imaging data were collected from a total of 62 subjects at time point 1, using a Siemens 1.5 Tesla scanner. These data included functional measurements of blood oxygenation level-dependent (BOLD) contrast and two structural scans: diffusion tensor imaging (DTI) and a T1-weighted sequence. From the DTI data we calculated fractional anisotropy (FA), which reflects white matter microstructure such as axonal ordering, axonal density, and degree of myelination. From the T1-weighted scan we calculated gray matter volume (GMV) using voxel-based morphometry by voxelwise multiplication of the segmented gray matter probability maps with the jacobian determinants from the nonlinear normalization to the MNI template. During the fMRI, subjects performed a WM and a control task. From these data an image of the contrast between the WM and the control task was calculated. The data corresponding to these three MRI modalities (DTI, GMV, BOLD) were used to train three SVR models (see below). Subjects who moved  $>4$  mm during scanning or had other significant artifacts were excluded from training the SVR model for the corresponding MRI modalities. Subjects who performed poorer than chance ( $n = 7$ ) on the fMRI WM task were excluded from the fMRI analysis due to expected poor WM BOLD signal. These exclusion criteria resulted in 59 subjects for training the DTI SVR, 55 for the GMV SVR, and 53 for the BOLD SVR. There were 46 subjects with three usable sequences that could be used when combining the MRI modalities in a multiple regression. When not including GMV, an additional 5 subjects could be included. The mean age of the excluded subjects for each modality was 9 ( $SD = 3.8$ ) years for fMRI, 12 ( $SD = 5.3$ ) years for DTI, and 7 ( $SD = 1.6$ ) for GMV. There were 16 subjects with one missing modality and 3 subjects with two missing modalities.

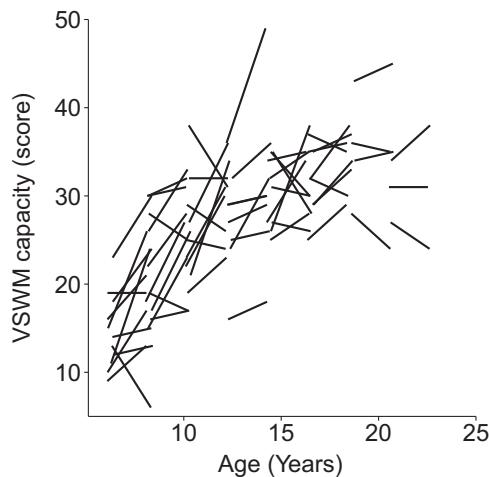
White matter was studied using a DTI sequence (field of view  $230 \times 230$  mm,  $128 \times 128$  grid,  $1 \text{ mm}^3$  voxels, 20 gradient directions). Eddy current and head motions were corrected with affine registration using FSL software ([www.fmrib.ox.ac.uk/fsl/](http://www.fmrib.ox.ac.uk/fsl/)). FA was calculated from the diffusion tensors for each voxel and used in subsequent SVR analyses.

Structural gray matter development was measured from T1-weighted spin echo images acquired with a 3D MPAGE sequence (field of view =  $256 \times 256$  mm,  $256 \times 256$  grid,  $1 \text{ mm}^3$  voxel size). The images were first segmented using SPM5 ([www.fil.ion.ucl.ac.uk/spm/software/spm5](http://www.fil.ion.ucl.ac.uk/spm/software/spm5)) and then normalized using Diffeomorphic Anatomical Registration using Exponential Lie algebra (DARTEL). Finally, an 8 mm Gaussian smoothing kernel was applied. The values of the normalized GMV images were then used as substrate for the SVR model.

For acquiring functional WM signal, the subjects performed a visuospatial WM task inside the MRI scanner. The task consisted of two sessions, each 4 min and 54 s. During each session the subjects were presented with 16 WM trials and 16 control trials. The WM trials consisted of a sequence of dots that were presented sequentially on a  $4 \times 4$  grid. The trials were divided equally between two WM loads, showing a sequence of 2 or 4 dots. Each dot was shown for 500 ms followed by a 500 ms interval. After the sequence of dots, there was a 1500 ms interval followed by a 3000 ms cue asking the subject to indicate whether a dot had appeared in a specified grid position at a specified position in the sequence. As an example, “2?” would appear in one of the grid squares

and the subjects had to press a button with the right index finger if the answer was true or with the middle finger if it was false. In the control condition, dots were shown in the corners of the grid and the cue was always the number 8 presented in a non-corner position. The order for WM and control trials was pseudorandomized. The E-Prime software was used together with an MRI-compatible visual system (Nordic, NeuroLab) to present the stimuli. During the stimuli T2\*-weighted functional images were acquired using a gradient echo Echo Planar Imaging pulse sequence (TR = 3000, TE = 50 ms, flip angle = 90°, 30 oblique slices, 4.5 mm slice thickness, 0.5 mm interslice distance,  $220 \times 220$  FOV,  $64 \times 64$  grid, voxel size  $3.44 \times 3.44 \times 4.5$  mm). Preprocessing of the fMRI data was performed with SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5>), implemented in Matlab (Mathworks). After normalization to MNI, space models were created using the canonical hemodynamic response function. Regressors were added for both WM and control loads. The WM regressors were defined as epochs starting at stimulus onset and lasting 8 or 10 s, respectively. Only correct trials were included in the model. A regressor for movement was also added in the model. Parametric contrast maps were created by contrasting both WM loads with the contrast condition. Beta values from the individual contrast maps were then used in subsequent SVR analysis.

**Support Vector Regression model and analysis.** A SVR model was trained using MRI data acquired at t1 to predict performance on the VSWM task at t2, 2 years later. To train the model, individual MRI volumes were initially represented as a vector. Subject vectors were then combined into one subject  $\times$  voxel matrix for each MRI modality. One SVR model was trained for each modality using these matrices. In the case of fMRI, all voxels with a mean positive  $\beta$  value in the contrast maps were included. No significance threshold was thus used. We used a  $v$ -SVR (Scholkopf et al., 1999) with  $v$  set to 1, since a linear kernel was used (see below). The other parameter of the SVR model  $C$  was set to the maximum  $t_2$  VSWM times the number of subjects, according to what is recommended in the literature (Chalimourda et al., 2004). Both of the parameters of the SVR model were therefore not fitted to the dataset but determined according to the usual criteria. Due to the large amount of features and, for machine learning purposes, the small number of subjects, a linear kernel was selected to minimize problems with overfitting. The generalization of the SVR model predictions was estimated using leave-one-out cross validation, which is commonly used when a completely independent testing sample is not available. One subject was therefore in turn left out when training the model. The testing was done for each subject using a model trained on the other subjects and performance was averaged over the subjects. We used general linear model analysis to combine the SVR models and to check whether they contributed significantly to predict t2 VSWM, when t1 VSWM and the inverse of age were also used as explanatory variables. The inverse of age rather than age was included because it has been shown to be a better predictor of cognitive development (Du-montel and Klingberg, 2012). This choice of age-related model was also justified with a separate analysis comparing it with a nonparametric regression as shown below. We also checked whether movement during the fMRI scans significantly contributed to predict t2 VSWM, when t1 VSWM was also added as an explanatory variable. Movement was calculated as the absolute average over the 6 movement parameters given as output as the image realignment of the data done with SPM5. This thus reflected the mean movement between two acquired volumes. This measure correlated with both age ( $r = -0.57$ ,  $p < 0.001$ ) and t1 VSWM ( $r = -0.59$ ,  $p < 0.001$ ), indicating that subjects move less in the scanner as they get older. However, the measure did not contain any predictive capacity of t2 VSWM ( $p > 0.9$ ), therefore not confounding the analysis. The average absolute movement parameter per age group was averaged over the standard 6 movement parameters, as follows: 5–7 years, mean = 0.033,  $SD = 0.015$ ; 7–9 years, mean = 0.026,  $SD = 0.018$ ; 9–11 years, mean = 0.024,  $SD = 0.009$ ; 11–13 years, mean = 0.022,  $SD = 0.012$ ; 13–15 years, mean = 0.016,  $SD = 0.008$ ; 15–17 years, mean = 0.010,  $SD = 0.003$ ; 17–19 years, mean = 0.010,  $SD = 0.001$ ; 19–21 years, mean = 0.009,  $SD = 0.004$ . It should be noted that this excludes linear relationships between in-scanner movement and the outcome. Complex, nonlinear distortions in the data might still be present, which is a limitation of this covariation. However, in the present study, we used linear



**Figure 1.** Development of VSTM in 62 healthy children and adolescents. Working memory tests were performed two times with an interval of 2 years.

kernels, minimizing the risk of only fitting residual nonlinear relationships.

To locate brain areas containing information about future working memory capacity, we conducted voxelwise general linear model analyses. Measurements of BOLD and FA were used separately as dependent variables, and t1 VSTM and imaging SVR prediction of t2 VSTM, or only t1 VSTM were used as independent or explanatory variables. Clusters were considered if they had a peak and voxel extent FWE corrected significance of  $p < 0.05$ . Single voxels significant at uncorrected  $p < 0.001$  were included to extend the clusters to represent a more regional signal for subsequent analysis (see below). This, however, did not change the cluster significance, which was the same when applying a voxelwise FWE-corrected  $p < 0.05$  as threshold.

Finally, we performed confirmatory general linear model analyses using t2 VSTM as dependent variable and t1 VSTM and the mean values of the clusters of BOLD or of FA as independent variables for each subject.

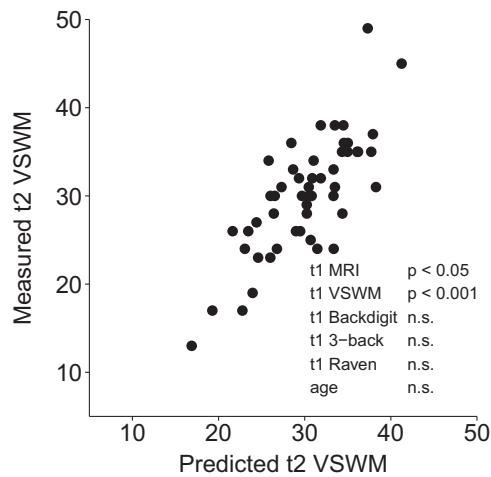
Coefficients for the SVR model itself were chosen to not be interpreted, because it is a multivariate statistic dependent on noise and signal in the other voxels, and should therefore be interpreted with caution (see Discussion).

Data modeling was done in Python (<http://www.python.org>). Libsvm (Chang and Lin, 2011) and Scikit-learn (Pedregosa et al., 2011) were used for the SVR analysis. Statistical analysis was conducted in the open source statistical computing environment R (<http://www.r-project.org/>).

## Results

To explore the predictive potential of MRI we used the functional and structural MRI data from subjects at t1 to train a SVR model using performance on the VSTM task at t2, 2 years later, as the dependent variable. We chose the VSTM task as the outcome measure due to its high predictive value regarding future mathematical achievement (Alloway, 2007). VSTM performance in children showed the expected increase with age (Fig. 1), and the well known high interindividual variance in performance and change over time, which highlights the need for subject-specific prediction of future development.

One SVR was trained for each imaging modality (DTI measuring white matter structure, GVM measuring gray matter volume and density, and BOLD measuring brain activity). All three imaging modalities provided significant predictions of future VSTM capacity, with the FA data from DTI having the strongest correlation ( $r = 0.59, p < 0.001$ ), followed by the BOLD data ( $r = 0.44, p < 0.001$ ) and the GVM data ( $r = 0.29, p < 0.05$ ). Combining all three imaging modalities in one regression model re-

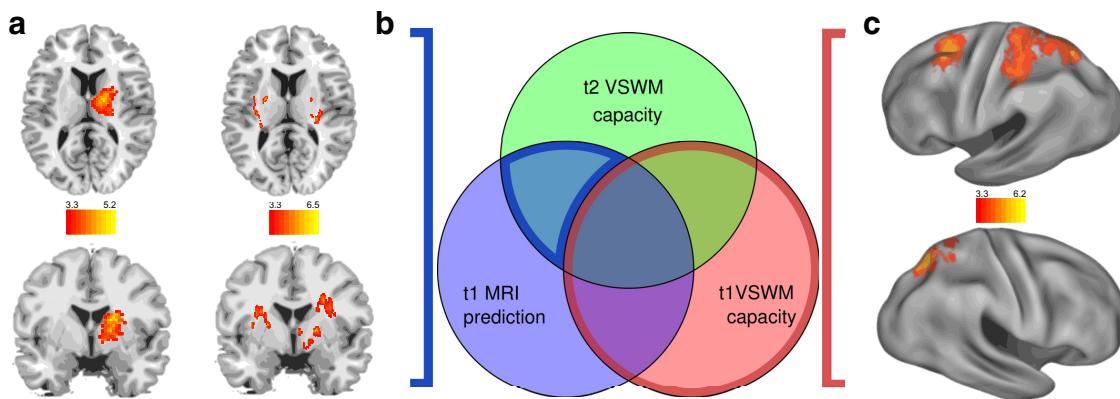


**Figure 2.** Prediction of t2 VSTM combining t1 VSTM and MRI prediction,  $r = 0.78$ ,  $p < 0.001$ ,  $n = 51$ .

sulted in a correlation of  $r = 0.64$  ( $p < 0.001$ ), with BOLD and DTI models independently contributing to the model (BOLD  $p < 0.05$ , DTI  $p < 0.01$ , and GVM  $p > 0.5$ ). These initial models were interpreted as representing both the effect of age and VSTM. Even though the neuroimaging SVR models were trained and evaluated using one dataset, the predictions obtained were unbiased and would not have the possibility to be overfitted. Due to the leave-one-out cross-validation, each prediction was derived from a model that was trained leaving the predicted subject out.

Next, we investigated whether the neuroimaging data could explain variance in future (t2) VSTM in addition to that explained by the cognitive test itself at baseline (t1). VSTM performance at t2 was the dependent variable and t1 neuroimaging prediction based on FA and BOLD, VSTM at t1, and the inverse function of chronological age ( $\text{age}^{-1}$ ) were independent variables in a multiple regression model. Age was included in this analysis to control for cognitively irrelevant, structural brain development. In this model the neuroimaging prediction could explain a significant proportion of the variance in future VSTM capacity ( $p < 0.05$ ). The entire model showed a correlation of  $r = 0.78$  ( $p < 0.001$ ) (Figure 2). In contrast, when the two other WM measures ( $n$ -back and digit-span) and a reasoning measure (Raven's Standardized Progressive Matrices) were included as independent variables instead of the neuroimaging measure, they were not able to explain additional variance ( $p > 0.4$ ). The significance of the neuroimaging prediction was not changed when the two WM measures and reasoning were included. We also addressed the hypothesis that the imaging data might be predicting t2 VSTM because it reflects subjects' movement, by checking whether movement during scanning was able to predict t2 VSTM when t1 VSTM was also included as a predictor. Movement was found to not contribute to prediction ( $p > 0.5$ ). Due to the wide age range of the subjects, we explored the interaction between predicted VSTM and age; however, no such interaction was found ( $p > 0.2$ ).

When the change in VSTM capacity (t2 minus t1) instead was used as the dependent variable, the same significant contribution of the predictors persisted. The neuroimaging model was thus a better predictor of the change of VSTM capacity over time than related WM tests together with the reasoning ability test, which is known to correlate highly with WM and general cognitive func-



**Figure 3.** **a**, BOLD correlates of future working memory capacity. **b**, FA correlates of future working memory capacity. **c**, BOLD correlates of current working memory capacity. Both **a** and **b** show unique variance after accounting for the effect of current working memory capacity for predicting future capacity. Voxel colors indicate *t* score. The diagram illustrates common and unique variance of t2 working memory capacity predicted by t1 VSWM and t1 MRI prediction. Voxels are significant at  $p < 0.001$  uncorrected. Clusters are significant at extent and peak FWE-corrected  $p < 0.05$ .

tioning (Crone et al., 2006). The VSWM still had a higher correlation between the two time points ( $r = 0.75$ ) than the correlation between the neuroimaging model and t2 VSWM ( $r = 0.64$ ). This shows that neuroimaging was able to explain change in VSWM capacity between t1 and t2, indicating that neuroimaging could predict future cognitive development.

The two-step regression described above may overestimate the contribution of the neuroimaging variable, because this variable is constructed to predict the same dependent variable (t2 VSWM). The severity of this bias should correlate with the number of variables in the first step, and so it is expected to be small since only two predictors are used here. The influence of the bias was explored by pairwise permutation of the DTI and fMRI predictors in the first regression step, thus keeping the relationship between the covariates while destroying their correlation with the dependent variable. The vector for the resulting model fit was entered into a regression model including the covariates outlined above. The *t*-statistic for this vector in the subsequent model was obtained over 100,000 permutations to obtain an empirical null distribution. The true statistic obtained with the non-permuted predictions was compared against this distribution. The model described above, where age, VSWM at t1, and two other WM scales as well as reasoning was included, resulted in  $p = 0.07$  for the combined imaging prediction, thus indicating a small difference compared with the standard normal distribution. When repeating the permutation test only including age as a covariate, the result was  $p = 0.01$  for the significance of the neuroimaging predictor, showing an age-independent significance of the imaging predictions. Thus, the bias induced by the two-step regression approach was small in our case.

To explore the generalization of the predictive model, a composite measure for WM was constructed using a structural equation model. The resulting model was used to calculate a composite WM score for t2. The correlation between WM composite scores for t1 and t2 was  $r = 0.81$  ( $p < 0.001$ ). The neuroimaging prediction correlated well with this t2 composite score ( $r = 0.60$ ,  $p < 0.001$ ) suggesting generalization. However, when covarying for age and t1 composite score, this did not remain significant ( $p = 0.17$ ), suggesting some limitation in this generalization.

The previous analysis showed that MRI could explain unique variance in future VSWM capacity that could not be predicted based on the cognitive tests. Next, we wanted to anatomically localize where this information was derived from. The coeffi-

cients, or weights, from the SVR model do not represent univariate statistical significance and may poorly reflect neural network structure (Michel et al., 2012; see Discussion). Therefore, we used a separate voxelwise GLM analysis for anatomical localization, rather than examining voxels with highly positive or negative coefficients from the SVR model. The signal for each voxel from the BOLD contrast images and the FA images was fitted separately in a multiple regression using t1 VSWM and imaging SVR prediction of t2 VSWM as covariates. In this way, we could localize signals that were correlated with the t2 prediction but independent of t1 VSWM. Significance for all univariate analyses was considered as a FWE-corrected  $p < 0.05$  for peak and cluster extent.

In the BOLD image, the SVR prediction of t2 VSWM was associated with a cluster encompassing the anterior part of thalamus, caudate nucleus, globus pallidus, and putamen in the right hemisphere with a maximum located at 10, -10, 8 (MNI XYZ, mm; Fig. 3a; FWE,  $p < 0.05$ ). For the FA images the prediction covariate was associated with FA values surrounding the basal ganglia bilaterally, in the centrum semiovale and internal capsule bilaterally (Fig. 3b). The FA correlation resulted in two large clusters with peaks at MNI XYZ: -22, -20, 2 mm and 19, -2, 5 mm, containing 5163 and 6160 voxels, respectively, and a smaller cluster containing 415 voxels at MNI XYZ: -42, -21, 26 mm was also found (FWE,  $p < 0.05$  for peak and cluster).

The variable t1 VSWM did not show any unique (not accounted for by the SVR prediction) neural correlates in BOLD or FA. In contrast, in a model where t1 VSWM was used as the only predictor, there was an association with the BOLD signal in parietal cortex bilaterally and premotor and prefrontal cortex in the left hemisphere (Fig. 3c), consistent with previous cross-sectional studies of VSWM development (Klingberg et al., 2002; Kwon et al., 2002; Crone et al., 2006; Scherf et al., 2006). Together, these results showed that signals localized to the thalamus, basal ganglia, and surrounding white matter predicted future WM, while current performance correlated with frontoparietal cortical activity.

The t1 SVR imaging prediction was derived from the t1 MRI images and was therefore not independent from the BOLD and FA measurements. As a confirmatory analysis, we therefore predicted t2 VSWM from the MRI clusters to ensure that the effects were not due to autocorrelation. Mean values from the clusters for BOLD and FA were each entered into a separate regression model predicting t2 VSWM while including t1 VSWM as a cova-

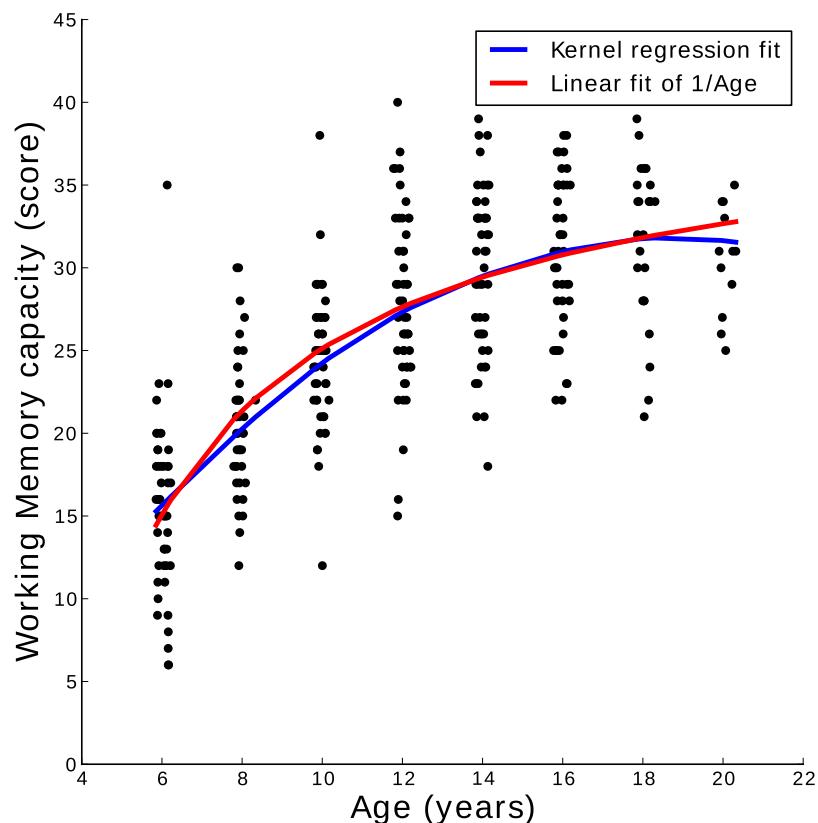
riate. Indeed, both BOLD and FA each had significant correlation coefficients in the model ( $p < 0.05$ ,  $p < 0.05$ ). This confirmed that predictive information about t2 VSWM was located in the thalamus, basal ganglia, and surrounding white matter. The partial correlation between t2 VSWM and BOLD and FA values, when t1 VSWM was accounted for, were 0.34 and 0.30, respectively. We report these values for completeness. Note, however, that the values are biased to higher values of correlation since the clusters were selected based on predicting t2 VSWM.

The correction for age was done using the inverse function of age. One possible problem with this choice would be if the true relationship between WM and age agreed with this function poorly. The correction for age would thus not capture the variance explained by this factor properly. To address this we compared the inverse fit with a nonparametric kernel regression. This regression fit a linear model that varies with age and will weigh the data points using a kernel. Data points are weighted based on their proximity in age. This gives rise to a nonlinear fit of which smoothness is determined by the kernel bandwidth. The kernel bandwidth is a parameter that has to be defined by the user. To obtain an optimized and unbiased fit, this parameter was set using a leave-one-out cross validation. We used a larger sample of 323 children from which the currently studied imaging sample was drawn. Both models gave a similar fit for the cognitive data. When the models were compared, no significance for superiority for either model was found ( $p > 0.1$  for  $\Delta AIC$ ). This shows that an observer-independent nonparametric fit for the age–WM relationship could not capture any variance not accounted for by the inverse fit. Visual inspection of the two functions support that both models are appropriate for the relationship (Fig. 4).

The main goal of the present study was to predict VSWM over development. As an extension of the present results, we tried to predict the residuals of the effect of age on VSWM. These residuals were obtained by the nonparametric regression described above. The same algorithm for the SVR was used. These predictions were not significant ( $p > 0.1$ ) for either DTI or fMRI. Thus, this extension to find an SVR pattern only corresponding to the age-independent variance was not possible with the current dataset. For further discussion, see below.

## Discussion

We showed that structural and functional imaging data could be used to predict the individual course of WM development in healthy children and adolescents. The predictive signal correlated most highly with BOLD activity in the thalamus and caudate nucleus, and FA in the surrounding white matter. The neuroimaging prediction suggests that there is unique MRI-derived information about future capacity which, in the present sample, could not be derived from either the VSWM task itself other WM tasks, a reasoning task, or age. In contrast, the WM performance at the time of scanning was correlated to activity in the frontal and



**Figure 4.** Curve fit to explore the relationship between age and WM. The curves were fit using the larger sample of cognitive measurements. Two functions, a linear fit for the inverse function of age and a nonparametric fit using kernel regression, were compared. The inverse function of age is transferred back into the non-inverse space for comparison. No statistical difference for the fit of the curve could be shown.

parietal cortex, consistent with previous cross-sectional studies of WM development in children (Klingberg et al., 2002; Kwon et al., 2002; Crone et al., 2006; Scherf et al., 2006; Geier et al., 2009).

Although we could show that neuroimaging data provided unique information, it is impossible to exclude that some other cognitive task, not included in this study, would be of similar predictive value, which is a limitation of the study. However, future performance of a test generally is most highly correlated to earlier performance on that very same test. The fact that neuroimaging provided unique information about future VSWM performance that could not be derived from earlier performance on the VSWM task itself is thus encouraging. Furthermore, if the neural basis for current and future capacity is partly separate, as these results suggest, this indicates a rationale why imaging could have a unique role in prediction. This could be compared with predictions of other biological systems such as how x-ray of the epiphyseal plate in adolescents can be used to predict further skeletal elongation.

An alternative explanation would be that the MRI only provided an additional measurement of VSWM, thus explaining additional variance in VSWM due to repeated measurement of the same cognitive system. The fact that cognitive tests and neuroimaging correlated with spatially separate systems does not, however, support this.

The statistical analysis was designed in two steps. The first step consisted of machine learning models predicting future VSWM capacity and allowed us to maximize the amount of information that could be extracted from the neuroimaging data. However, no univariate significance test on the voxels was included in this

analysis. Therefore, voxels in the model could not be said to individually contain information about future VSWM. Only the predictions obtained from the model were tested for significance. High or low voxel weights for single voxels in the SVR can be due to noise, and even clusters of high- or low-weighted voxels can be due to a Gaussian distribution of the noise in MRI data. Therefore, areas with redundant information can show a spatial distribution of the coefficients mimicking a biological pattern even though it is resembling only the spatial distribution of noise. This is a fact sometimes overlooked in neuroimaging studies with multivariate analyses but is of high importance for interpretation of the results. Many multivariate methods, including SVR, have been shown to poorly reflect the underlying ground truth activity pattern in the individual voxel weights (Michel et al., 2012), stressing this point. We therefore chose not to use the SVR model for the second objective of the analysis, to spatially localize the information extracted with the machine learning models. Instead we used the prediction obtained from the SVR model as the independent variable in a subsequent mass-univariate GLM. Using this two-step approach, we could use both the high predictive performance of machine learning models and the more reliable brain-mapping abilities of univariate statistical models such as GLM. Thus, we used two state of the art statistical methods that are commonly used in a number of fields for accomplishing our goal. Future research in the interface between neuroimaging and machine learning has yet to determine the best suited models for this purpose.

Here we aimed to compare the information content of neuroimaging data with that of cognitive measurements. One problem when comparing univariate predictors such as cognitive measurements with highly multivariate predictors such as MRI data is that it is very difficult to exclude a possible lack of predictive capacity due to suboptimally modeled data. Difficulties in excluding the confounding of nonlinear effects from multivariate datasets is a limitation of neuroimaging in general, but this study in particular, due to the aim of finding unique information rather than covarying for linear confounders.

To explore the change in working memory over development, we used a longitudinal sample. Since we were interested in the developmental aspects of WM, we chose to not remove the effect of age on working memory before training the SVR model. However, we tried to extend the results by only predicting the developmentally unrelated variance in WM, which is a separate question since the majority of WM variance is dependent on age. We did this by trying to predict the residuals of the kernel regression relationship for age on WM. This prediction was not significant. We see two possible explanations for this. First, the residuals of WM will be more noisy. The test-retest correlation for the residuals was  $r = 0.58$ , while VSWM had a test-retest correlation of  $r = 0.75$  in the larger behavioral sample. Predicting a noisy variable is expected to reduce the SVR performance significantly. Second, due to the sample size, we were limited to linear functions in this study. The relation between age-corrected or centered working memory and brain structure or activity is likely to be much weaker and eventually more complex than that associated with age. Such a subtle relation might be difficult to predict with a linear kernel as used in this study.

The basal ganglia and thalamus are tightly linked with WM-related cortical areas in the frontal and parietal lobe through the basal ganglia-thalamo-cortical loop (Alexander et al., 1986; Alexander and Crutcher, 1990; Aosaki et al., 1995). The basal ganglia-thalamo-cortical loop is important for implicit learning and habit formation, a system separate from the medial temporal lobe sys-

tem underlying encoding into declarative long-term memory (Packard and Knowlton, 2002; Graybiel, 2008). In a study of associative learning using a delayed-response task in primates, basal ganglia activity changes preceded the slower change in prefrontal cortex, where the latter matched the changes in behavior (Pasupathy and Miller, 2005). This finding supports the hypothesis that rewarded associations are first identified by the basal ganglia which, then, via the basal ganglia-cortical loops, promotes the slower Hebbian learning of the frontal cortex which determines the behavioral performance. In a recent study (Darki and Klingberg, 2014) the striato-frontal pathways were identified by tract tracing in a developmental sample overlapping with the sample in the current study. Darki and Klingberg (2014) here found that fractional anisotropy in the striato-frontal pathway predicted WM 2 years later, a result which confirms the present findings and extends it by higher anatomical specificity. This role in implicit learning is also consistent with a study of young adults learning to play a video game, where it was found that caudate and putamen size predicted the rate of learning over 20 sessions (Erickson et al., 2010). In contrast, no correlation was found between learning and the size of the hippocampus or ventral striatum (nucleus accumbens), emphasizing the difference between declarative and implicit learning.

The basal ganglia and thalamus have not been specifically implicated in the development of WM. Although the caudate nucleus is activated in children performing WM tasks (Klingberg et al., 2002; Ziermans et al., 2012), the activity is not correlated with current capacity in these cross-sectional analyzes, in contrast to the activity in parietal and frontal cortices. However, behavioral studies suggest that WM can be improved by training (Klingberg, 2010). Two fMRI studies of training-induced improvements in WM implicated both thalamus and caudate nucleus (Olesen et al., 2003; Dahlin et al., 2008). In one training study, the amount of striatal activity predicted the amount of improvement seen after training (Dahlin et al., 2008). Working memory training results in changes of D2 occupancy (Bäckman et al., 2011), and the changes in capacity after training are associated with changes in cortical D1 density (McNab et al., 2009). These results provide an indirect association between training and development. If the circuitry for implicit learning and cognitive training is also relevant for cognitive development, it suggests that experience and learning are important for childhood development of executive functions, which would be a new perspective on executive development.

In summary the results from the current study suggest that future cognitive development can be predicted from anatomical and functional MRI signals above and beyond that explained by cognitive tests. The specific predictive ability of neuroimaging seems to be due to the fact that it is possible to distinguish a subcortical system predicting future cognitive development from a cortical system associated with current cognitive capacity. This has wide implications for understanding the neural mechanisms of cognitive development. It also supports the possibility that cognitive development can be predicted using MRI in a way not assessable by cognitive tests alone.

## References

- Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci* 13:266–271. CrossRef Medline
- Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357–381. CrossRef Medline

- Alloway T (2007) Automated working memory assessment (AWMA). London: Harcourt Assessment.
- Aosaki T, Kimura M, Graybiel AM (1995) Temporal and spatial characteristics of tonically active neurons of the primate's striatum. *J Neurophysiol* 73:1234–1252. [Medline](#)
- Bäckman L, Nyberg L, Soveri A, Johansson J, Andersson M, Dahlin E, Neely AS, Virta J, Laine M, Rinne JO (2011) Effects of working-memory training on striatal dopamine release. *Science* 333:718.
- Chalimourda A, Schölkopf B, Smola AJ (2004) Experimentally optimal  $v$  in support vector regression for different noise models and parameter settings. *Neural Networks* 17:127–141. [CrossRef](#) [Medline](#)
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27. [CrossRef](#)
- Crone EA, Wendelken C, Donohue S, van Leijenhorst L, Bunge SA (2006) Neurocognitive development of the ability to manipulate information in working memory. *Proc Natl Acad Sci U S A* 103:9315–9320. [CrossRef](#) [Medline](#)
- Dahlin E, Neely AS, Larsson A, Bäckman L, Nyberg L (2008) Transfer of learning after updating training mediated by the striatum. *Science* 320: 1510–1512. [CrossRef](#) [Medline](#)
- Darki F, Klingberg T (2014) The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. *Cereb Cortex*. In press.
- Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczkó E, Coalson RS, Pruett JR Jr, Barch DM, Petersen SE, Schlaggar BL (2010) Prediction of individual brain maturity using fMRI. *Science* 329: 1358–1361. [CrossRef](#) [Medline](#)
- Dumantheil I, Klingberg T (2012) Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cereb Cortex* 22:1078–1085. [CrossRef](#) [Medline](#)
- Erickson KI, Boot WR, Basak C, Neider MB, Prakash RS, Voss MW, Graybiel AM, Simons DJ, Fabiani M, Gratton G, Kramer AF (2010) Striatal volume predicts level of video game skill acquisition. *Cereb Cortex* 20:2522–2530. [CrossRef](#) [Medline](#)
- Gabrieli JD (2009) Dyslexia: a new synergy between education and cognitive neuroscience. *Science* 325:280–283. [CrossRef](#) [Medline](#)
- Gathercole SE, Pickering SJ, Knight C, Stegmann Z (2004) Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Appl Cogn Psychol* 18:1–16. [CrossRef](#)
- Geier CF, Garver K, Terwilliger R, Luna B (2009) Development of working memory maintenance. *J Neurophysiol* 101:84–99. [Medline](#)
- Graybiel AM (2008) Habits, rituals, and the evaluative brain. *Annu Rev Neurosci* 31:359–387. [CrossRef](#) [Medline](#)
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. [CrossRef](#) [Medline](#)
- Klingberg T (2010) Training and plasticity of working memory. *Trends Cogn Sci* 14:317–324. [CrossRef](#) [Medline](#)
- Klingberg T, Forssberg H, Westerberg H (2002) Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *J Cogn Neurosci* 14:1–10. [CrossRef](#) [Medline](#)
- Kwon H, Reiss AL, Menon V (2002) Neural basis of protracted developmental changes in visuo-spatial working memory. *Proc Natl Acad Sci U S A* 99:13336–13341. [CrossRef](#) [Medline](#)
- McNab F, Varrone A, Farde L, Jucaite A, Bystritsky P, Forssberg H, Klingberg T (2009) Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323:800–802. [CrossRef](#) [Medline](#)
- Michel V, Gramfort A, Varoquaux G, Eger E, Keribin C, Thirion B (2012) A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recogn* 45:2041–2049. [CrossRef](#)
- Olesen PJ, Westerberg H, Klingberg T (2003) Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci* 7:75–79. [CrossRef](#) [Medline](#)
- Packard MG, Knowlton BJ (2002) Learning and memory functions of the basal ganglia. *Annu Rev Neurosci* 25:563–593. [CrossRef](#) [Medline](#)
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876. [CrossRef](#) [Medline](#)
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. *J Machine Learn Res* 12:2825–2830.
- Raven J, Raven JC, Court JH (1998) Manual for Raven's progressive matrices and vocabulary scales. Oxford: Oxford Psychologists.
- Scherf KS, Sweeney JA, Luna B (2006) Brain basis of developmental change in visuospatial working memory. *J Cogn Neurosci* 18:1045–1058. [CrossRef](#) [Medline](#)
- Scholkopf B, Bartlett P, Smola A, Williamson R (1999) Shrinking the tube: a new support vector regression algorithm. *Adv Neural Inform Process Syst* 11:330–336.
- Söderqvist S, McNab F, Peyrard-Janvid M, Matsson H, Humphreys K, Kere J, Klingberg T (2010) The SNAP25 gene is linked to working memory capacity and maturation of the posterior cingulate cortex during childhood. *Biol Psychiatry* 68:1120–1125. [CrossRef](#) [Medline](#)
- Ziermans T, Dumantheil I, Roggeman C, Peyrard-Janvid M, Matsson H, Kere J, Klingberg T (2012) Working memory brain activity and capacity link MAOA polymorphism to aggressive behavior during development. *Translational Psychiatry* 2:e85. [CrossRef](#) [Medline](#)