



ELSEVIER

## Developmental Cognitive Neuroscience

journal homepage: <http://www.elsevier.com/locate/dcn>

## Accuracy of episodic autobiographical memory in children with early thyroid hormone deficiency using a staged event

Karen A. Willoughby<sup>a,b</sup>, Mary Pat McAndrews<sup>b,d</sup>, Joanne F. Rovet<sup>a,b,c,\*</sup><sup>a</sup> Neuroscience and Mental Health Research Program, The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada<sup>b</sup> Department of Psychology, University of Toronto, 100 St. George Street, 4th Floor, Sidney Smith Hall, Toronto, ON, Canada<sup>c</sup> Department of Pediatrics, University of Toronto, The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada<sup>d</sup> Krembil Neuroscience Centre, Toronto Western Hospital & Research Institute, 399 Bathurst Street, Toronto, ON, Canada

## ARTICLE INFO

## Article history:

Received 12 July 2013

Received in revised form

28 November 2013

Accepted 20 December 2013

## Keywords:

Thyroid hormone

Hippocampus

Autobiographical memory

Episodic memory

Accuracy

Staged event

## ABSTRACT

Autobiographical memory (AM) is a highly constructive cognitive process that often contains memory errors. No study has specifically examined AM accuracy in children with abnormal development of the hippocampus, a crucial brain region for AM retrieval. Thus, the present study investigated AM accuracy in 68 typically and atypically developing children using a staged autobiographical event, the Children's Autobiographical Interview, and structural magnetic resonance imaging. The atypically developing group consisted of 17 children (HYPO) exposed during gestation to insufficient maternal thyroid hormone (TH), a critical substrate for hippocampal development, and 25 children with congenital hypothyroidism (CH), who were compared to 26 controls. Groups differed significantly in the number of accurate episodic details recalled and proportion accuracy scores, with controls having more accurate recollections of the staged event than both TH-deficient groups. Total hippocampal volumes and anterior hippocampal volumes were positively correlated with proportion accuracy scores, but not total accurate episodic details, in HYPO and CH. In addition, greater severity of TH deficiency predicted lower proportion accuracy scores in both HYPO and CH. Overall, these results indicate that children with early TH deficiency have deficits in AM accuracy and that the anterior hippocampus may play a particularly important role in accurate AM retrieval.

© 2013 The Authors. Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

Episodic autobiographical memory (AM), or the recollection of specific past events, is a highly re-constructive

cognitive process involving the integration of many different types of episodic details (Svoboda et al., 2006). The majority of AM studies have focused on measuring the total number of episodic details recalled about past events, with greater details signaling better episodic AM performance (e.g., Levine et al., 2002; St-Laurent et al., 2009; Willoughby et al., 2012, 2013a,b). Although this method has many advantages, it may not provide the most sensitive measure of children's episodic AM if it does not involve verifying original event details or controlling for total verbal output (Koriat and Goldsmith, 1994). For example, some children may report a greater number of episodic details about past events than others simply because they have more developed language abilities. Other children may report inaccurate episodic details during memory

\* Corresponding author at: Department of Psychology, The Hospital for Sick Children, 555 University Avenue, Toronto ON, Canada M5G1X8.

Tel.: +1 416 813 8283; fax: +1 416 813 8839.

E-mail addresses: [karen.willoughby@utoronto.ca](mailto:karen.willoughby@utoronto.ca) (K.A. Willoughby), [mary.mcandrews@uhn.ca](mailto:mary.mcandrews@uhn.ca) (M.P. McAndrews), [joanne.rovet@sickkids.ca](mailto:joanne.rovet@sickkids.ca), [jrovet@sympatico.ca](mailto:jrovet@sympatico.ca) (J.F. Rovet).

interviews because they may feel pressured to provide a response when asked probing questions (Roebbers and Schneider, 2000). It is well-known that AMs often contain errors and are not completely accurate representations of past events (Hyman and Loftus, 1998). Accordingly, several researchers have suggested that measures of AM that both enable scoring for accuracy and control for total verbal output may provide a more sensitive measure of children's episodic AM ability than the total number of episodic details recalled about past events (e.g., Koriat and Goldsmith, 1994).

One promising technique for assessing AM accuracy is the use of staged events (i.e., demonstrations/activities where the content is controlled across participants). For example, Bruck et al. (2007) had children participate in a staged magic show and later assessed the accuracy of their recollections of this event. Staged events are beneficial because they allow for independent verification of original event details and reduce the variability across participants' recollections because all participants were exposed to the same event. Studies that have previously used staged events to examine AM accuracy in clinical pediatric populations, such as children with autism or intellectual disability, have shown that after controlling for total verbal output, these groups provide proportionally less accurate recollections of past events than controls (Agnew and Powell, 2004; Bruck et al., 2007). Thus, staged events may provide an important avenue for investigating AM accuracy in children who exhibit abnormal development of the hippocampus, a critical brain structure for episodic AM (Eldridge et al., 2000; Moscovitch, 2008).

Recent advances in neuroimaging have also prompted significant interest in AM accuracy by specifically examining whether the hippocampus is preferentially activated during accurate/successful episodic AM retrieval. Although the prefrontal cortex plays an essential role in monitoring and verifying retrieved information, research also indicates that the hippocampus is important for accurate AM retrieval because it mediates retrieval processes supported by the prefrontal cortex and reintegrates episodic details linked together in a memory trace (Addis et al., 2007; Brassens et al., 2006; Cabeza et al., 2004; Kircher et al., 2008; Skinner et al., 2009). For example, Cabeza et al. (2004) examined AM retrieval accuracy using a novel photo paradigm that controlled for the content of participants' AMs. These researchers found significant activation within the hippocampus, medial prefrontal cortex, and parahippocampal gyrus during *correct* recognition trials, which may signify hippocampal involvement in accurate AM retrieval (Cabeza et al., 2004). Other neuroimaging studies examining episodic recognition memory have found similar results, in that retrieval success (or accuracy) was associated with robust activity in the hippocampus and medial temporal lobe (e.g., Eldridge et al., 2000; Montaldi et al., 2001; Stark and Squire, 2000). In addition, several studies have reported that the anterior hippocampus appears to be especially important for successful episodic encoding and retrieval (Kircher et al., 2008; Sperling et al., 2003). Given this evidence, it is possible that children with abnormal hippocampal development may have difficulty recalling accurate episodic AM details from a staged event.

However, no study to date has used a staged event to examine this issue.

The main aim of the present study was to investigate AM accuracy in children exposed to insufficient thyroid hormone (TH) during gestation, which has been shown to have to significant adverse effects on hippocampal structure and function (Gilbert and Paczowski, 2003; Lavado-Autric et al., 2003). In particular, AM accuracy was assessed in two populations with early TH deficiency: (a) children of treated hypothyroid women (HYPO), who received insufficient maternal TH throughout gestation, and (b) children with early-treated congenital hypothyroidism (CH), who experienced late prenatal and early postnatal TH deficiency due to fetal thyroid gland dysfunction (Zoeller and Rovet, 2004). Previous studies have shown that both of these populations exhibit weaknesses in episodic AM (Willoughby et al., 2013a,b) and reduced hippocampal volumes relative to controls (Wheeler et al., 2011; Willoughby, 2011; Willoughby et al., 2013a,b). To assess AM accuracy, a staged autobiographical event was developed and participants' recollections of this event were assessed using the Children's Autobiographical Interview (CAI; Willoughby et al., 2012). The main variable of interest was proportion accuracy scores (i.e., percentage of total episodic details recalled that were accurate), in order to control for total verbal output and any possible group differences in language ability. Based on the results of previous studies examining AM accuracy in other clinical pediatric populations (Agnew and Powell, 2004; Bruck et al., 2007), it was expected that both CH and HYPO groups would exhibit proportionally less accurate recollections of the staged event than controls.

Given that the hippocampus is a critical structure for episodic AM, a secondary aim was to determine whether total hippocampal volumes predicted proportion accuracy scores and/or total episodic details reported. Based on suggestions that proportion accuracy scores may provide a more sensitive measure of children's episodic AM and hippocampal function than the total number of episodic details reported (e.g., Koriat and Goldsmith, 1994), we expected to find significant relations only between proportion accuracy scores and hippocampal volumes. We hypothesized that smaller hippocampal volumes would predict lower proportion accuracy scores. Finally, we also examined the relation between proportion accuracy scores and severity of TH deficiency (i.e., elevated maternal or neonatal TSH values) in HYPO and CH.

## 2. Material and methods

### 2.1. Participants

Participants were 68 children (46% male), aged 10–14 years ( $M = 10.71$ ,  $SD = 1.17$ ), including 25 CH, 17 HYPO, and 26 typically developing controls. All CH participants except two were identified at birth through the Ontario newborn screening program, which uses a thyroid stimulating hormone (TSH) test to screen for hypothyroidism (see Table 1 for TSH values in CH). The exceptions were two males born outside of Canada, one diagnosed at 14 days of age and one with an ectopic gland, who was not formally diagnosed

**Table 1**

Group means (range) for TH values for CH and mothers of HYPO.

	CH (N = 25)	HYPO (N = 17)
Trimester 1 TSH value (mU/L) <sup>a</sup> (Normal range during pregnancy < 2.5 mU/L)	–	9.76 (.10–28.11)
Trimester 2 TSH value (mU/L) <sup>b</sup>	–	7.38 (.45–38.50)
Trimester 3 TSH value (mU/L) <sup>c</sup>	–	4.90 (.10–17.60)
TSH value at birth (mU/L) <sup>d</sup> (Normal range = .4–4.1 mU/L)	135.00 (20.00–250.00)	4.54 (1.50–12.60)
TSH value at diagnosis (mU/L) <sup>e</sup>	352.75 (9.10–1172.80)	–
Total T4 value at diagnosis (nmol/L) <sup>f</sup> (Normal range = 50–150 nmol/L)	72.21 (5.00–229.00)	–
Free T4 value at diagnosis (pmol/L) <sup>g</sup> (normal range = 10–23 pmol/L)	8.93 (1.40–21.30)	–
Age at diagnosis (in days) <sup>h</sup>	42.75 (5.00–730.00)	–

Note: TSH = thyroid stimulating hormone, higher TSH levels indicate more severe TH. Deficiency, T4 = thyroxine.

<sup>a</sup> 4 missing cases.

<sup>b</sup> 3 missing cases.

<sup>c</sup> 3 missing cases.

<sup>d</sup> 6 missing CH cases and 6 missing HYPO cases.

<sup>e</sup> 2 missing cases.

<sup>f</sup> 5 missing cases.

<sup>g</sup> 2 missing cases.

<sup>h</sup> 1 missing case.

until 2 years of age but did not exhibit dramatically elevated TSH value at diagnosis (i.e., his TSH value at diagnosis was 10 mU/L). All CH participants had elevated TSH values at birth or at diagnosis. The entire HYPO group belonged to a cohort recruited at birth and followed throughout childhood at the Hospital for Sick Children (SickKids). Mothers of the HYPO group were originally identified through the SickKids' Motherisk program or through local endocrinologists. All mothers of the HYPO group were hypothyroid during pregnancy and received synthetic TH treatment. Maternal TSH values were the highest during the first trimester, most likely due to inadequate TH treatment during the first half of gestation (see Table 1 for maternal TSH values). Finally, all controls were recruited from a list of past participants, 25 of whom were from a longitudinal cohort followed at SickKids.

Exclusionary criteria for all participants were exposure to alcohol or other teratogens during pregnancy, a head injury, neuroradiological abnormality, chronic medical condition, psychiatric diagnosis, and an IQ score below 80. Additional exclusionary criteria for controls were an identified learning disability and a diagnosis of ADHD. Four participants did not complete the structural imaging portion of the study because they did not want to undergo scanning (2 CH, 1 HYPO), or moved so excessively during the scan that their data had to be discarded (1 CH). All participants were fluent or native English speakers.

## 2.2. General procedure

All parents provided written consent for participation in this study while participants also provided verbal assent. Participants were individually tested in the Psychology Department at SickKids over the course of two days, separated by four and a half months ( $M = 4.60$ ;  $SD = 2.37$ ). On Day 1, participants took part in a clinical neuropsychological assessment of their intelligence and general memory ability (not assessed presently), while parents completed several questionnaires. During their

lunch break, all participants also took part individually in a 45-minute staged event at SickKids. At the start of the staged event, participants were given detailed instructions about the staged event (see Willoughby, 2011, Appendix J) as well as 10 dollars to purchase a meal for lunch. During the staged event, participants received a tour of the hospital, answered a set of standard questions (e.g., regarding the date, current location in the hospital, the weather outside, etc.), and performed a series of tasks (e.g., played 20 questions, counted footprints on the floor, identified objects in a wall display, bought a lunch, watched a video, smelt a fragrant plant). Effort was made to ensure that the staged event was experienced identically across all participants by having the examiner use scripted statements (see Willoughby, 2011, Appendix K, for a detailed script of the staged event). The staged event was designed so that it contained enough unique event happenings, time, place, and visual/sensory/perceptual details to be comparable to naturally occurring autobiographical events.

On Day 2, participants were interviewed about the staged event using the Children's Autobiographical Interview (CAI; for a full description of the CAI, see Willoughby et al., 2012). One examiner (K.W.) conducted 99% of the staged events and interviews, while a second trained examiner conducted the staged event and interview for one CH participant. Immediately after the CAI, participants underwent structural MRI scanning in a 1.5 Tesla Signa GE scanner, during which they viewed movies via MRI-compatible goggles in the Diagnostic Imaging Unit at SickKids. A neuroradiologist examined all MRI images for neuroradiological abnormalities and sent written reports to participants' family physicians. At the end of Day 2, participants received a movie gift certificate, certificate of participation, and a CD containing several pictures of their own brain. Parents received compensation for transportation as well as a report describing their child's performance on the neuropsychological assessment from Day 1. All procedures were approved by the Research Ethics Board of SickKids and the Office of Research Ethics at the University of Toronto.

### 2.3. Demographic information and assessment of intelligence

Demographic information was obtained from all parents using a questionnaire, which yielded information about the child's age, sex, ethnicity, primary language, overall health, psychiatric health, the parents' marital status, and family socioeconomic status (SES; Hollingshead, 1975). Intelligence was evaluated using the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), which yields an estimate of Full Scale IQ.

### 2.4. Early TSH levels and TH treatment values for CH and HYPO

Thyroid stimulating hormone (TSH) is the gold standard for diagnosing hypothyroidism, as it is the most sensitive and reliable measure of thyroid abnormalities (Glinioer and Spencer, 2010).

For CH participants, information on newborn TSH values was obtained from medical health records at SickKids (see Table 1). For HYPO participants, maternal TSH values during pregnancy were obtained from the participants' family physicians (see Table 1).

### 2.5. Children's autobiographical interview

Participants' recollections of the staged event were assessed using the CAI (Willoughby et al., 2012). Additional questions about the staged event were added in the Specific Probe phase of the CAI (see Willoughby, 2011, Appendix L, for a list of these additional questions). Importantly, these additional questions were designed to limit suggestibility and required participants to provide an actual detail rather than a yes/no response. Participants completed all three phases of the CAI (i.e., free recall, general probe, and specific probe) for the staged event and provided a self-report rating of their ability to visualize the staged event at the time of recollection.

### 2.6. Scoring protocol for the children's autobiographical interview

For the present study, the standard CAI scoring protocol was used (see Levine et al., 2002 for a more detailed description), except that: (a) details recalled across the three phases of the CAI (free recall, general probe, and specific probe) were collapsed instead of assessing each phase separately, (b) only episodic/internal details (not non-episodic/external details) were scored and were summed to provide a *total episodic details* composite score, (c) details from the emotion/thought subcategory of episodic details were excluded from all analyses assessing accuracy because they could not be objectively deemed accurate or inaccurate, (d) details from each episodic subcategory (i.e., event, place, time, and perceptual details) were further subdivided into *inaccurate* and *accurate* detail subcategories (see below for a description), and (e) the accurate perceptual detail subcategory was further subdivided into accurate *visual* and accurate *non-visual* detail subcategories

(see Willoughby et al., 2013a,b for further description of visual and non-visual episodic detail criteria).

*Inaccurate details* included any reported episodic detail that was deemed incorrect based on detailed records from each participant's staged event. The majority of inaccurate details were either approximations (i.e., imprecise details) or confabulations (i.e., completely false details; see Table 2a for examples). A *total inaccurate details* composite score was created by summing all of the details from each inaccurate episodic detail subcategory (i.e., event, place, time, and perceptual). *Accurate details* referred to any reported episodic detail that was deemed correct based on detailed records of each participant's staged event (see Table 2b for examples). A *total accurate details* composite score was created by summing all of the details from each accurate episodic detail subcategory (i.e., event, place, time, and perceptual details).

Finally, in order to assess AM accuracy after controlling for total verbal output, a *proportion accuracy* score was created by dividing the total number of accurate episodic details recalled by the total number of episodic details recalled (i.e., excluding emotion/thought details). All memories were scored by K.W. and 46% of the memories selected at random within each group (10 CH, 10 HYPO, 14 CON) were re-scored by a second extensively trained scorer for reliability purposes. Both scorers were blind to participants' group, age, and sex. Inter-rater reliability was determined using intra-class correlations (one-way random effects model; McGraw and Wong, 1996). Coefficients for total inaccurate details, total accurate details, and proportion accuracy scores were .90, .90, and .85, respectively.

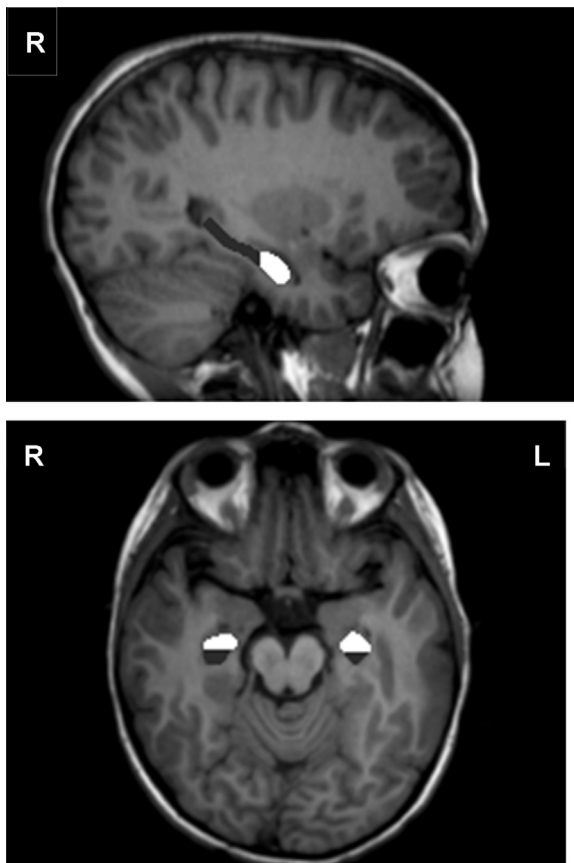
### 2.7. Structural MRI acquisition, processing, and hippocampal volumes

Structural images used for volumetric analysis were collected using a 3D FSPGR (IR prep Zip X2) T1-weighted series in the axial plane (TR=10.37 ms, TE=4.26 ms, TI=400 ms, flip angle=20°, field of view=240 mm, slice thickness=1.50 mm, no gap). Post-processing of MRI images involved extracting brain images from the skull and cerebral spinal fluid and normalizing to standard space using anterior and posterior commissure alignment. Intracranial volumes were determined using SPM5 software (Statistical Parametric Mapping 5; Wellcome Department of Imaging Neuroscience). One trained analyst (K.W.), masked to group status, manually traced both left and right hippocampal volumes on a slice-by-slice basis with a trackball-driven cursor using ANALYZE 9.0 software.

Designation of the hippocampal region included the CA fields, dentate gyrus, and the subiculum. Hippocampal boundaries were outlined from anterior to posterior in coronal images, such that designation began at the rostral end when the hippocampal head first appeared below the amygdala and ended when the crura of the fornices departed from the hippocampal tail at the caudal end (Pruessner et al., 2000). In addition, each hippocampus was further subdivided into anterior (i.e., hippocampal head) and posterior (i.e., hippocampal body and tail) segments along the coronal axis using the uncus as an unambiguous dividing marker (see Fig. 1), following procedures

**Table 2**  
Descriptions and examples of inaccurate and accurate episodic details of the staged.

Category	Description	Examples from Staged Event
(a) Inaccurate episodic detail subcategories and types of errors		
(i) Event		
Approximation	Imprecise event detail	"I had cheese pizza for lunch" (correct = pepperoni)
Confabulation	Completely false event detail	"It was raining outside" (correct = sunny)
(ii) Place		
Approximation	Imprecise place detail	"We started on the 5th floor" (correct = 6th floor)
Confabulation	Completely false place detail	"I ate lunch in the cafeteria" (correct = lunch room)
(iii) Time		
Approximation	Imprecise time detail	"It was on August 15th" (correct = 16th)
Confabulation	Completely false time detail	"The lunch started at 3pm" (correct = 12pm)
(ii) Perceptual		
Approximation	Imprecise perceptual detail	"You wore a purple sweater" (instead of pink)
Confabulation	Completely false perceptual detail	"I saw needles in the display" (incorrect - coins)
(b) Accurate episodic detail subcategories		
(i) Event		
	Accurate description of actions, people involved, clothing, and the weather	"I counted the footprints"
(ii) Place		
	Accurate descriptions of location (province, city, street, building, room, part of room)	"We started on the 6th floor"
(iii) Time		
	Accurate descriptions of year, season, month, date, day of week, and time of day	"It was in January"
(iv) Perceptual		
	Accurate descriptions of sensory and visual details, body position, orientation in space, duration of event	"I was sitting on the right side of the table"



**Fig. 1.** Manual tracings of anterior (white) and posterior (dark gray) subregions of the left and right hippocampus using ANALYZE 9.0 SoftwareR.

outlined in [Weiss et al. \(2005\)](#). To reduce manual tracing errors, coronal images were magnified three times and both sagittal and transverse images were also used to verify hippocampal boundaries. A second extensively trained image analyst traced 36% of right and left hippocampal volumes, which were selected at random within each group (6 CH, 7 HYPO, 8 CON). An intra-class correlation analysis using a one-way random effects model ([McGraw and Wong, 1996](#)) indicated coefficients of .84 (Cronbach's  $\alpha = .93$ ) for the right hippocampus and .88 (Cronbach's  $\alpha = .95$ ) for the left hippocampus.

### 2.8. Data analysis

Regarding demographic variables, chi-squared tests were used to compare groups on sex distribution, and ANOVAs were used to compare groups on age (measured in years), SES, and IQ scores. For data from the CAI, group differences in proportion accuracy scores were examined using ANCOVA, with age, sex, and retention interval as covariates. Although retention interval did not differ significantly across groups ( $p = .314$ ; see [Table 3](#) for group means), it was used as a covariate because it was significantly correlated with total episodic details;  $r = -.377$ ,  $p = .002$ , total accurate details;  $r = -.322$ ,  $p = .007$ , and all of the accurate detail subcategories ( $ps < .05$ ).

Group differences in total inaccurate and total accurate details were examined using a mixed-factor ANCOVA, with detail category (inaccurate/accurate) as the repeated measure, group as the between subjects variable, and age, sex, and retention interval as covariates. Group differences in inaccurate and accurate details subcategories were examined using separate mixed-factor ANCOVAs with the Greenhouse–Geisser correction applied. Detail

**Table 3**  
Frequency scores, group means (SD) for demographic and hippocampal volume data.

	Controls (N = 26)	CH (N = 25) <sup>a</sup>	HYPO (N = 17) <sup>b</sup>	p-Value <sup>c</sup>
Learning disability	0	2	0	.176
Diagnosis of ADHD	0	2	1	.370
Sex (number of males)	9	11	11	.155
Socioeconomic status	52.35 (10.44)	48.12 (9.68)	50.29 (10.41)	.338
Age	10.46 (.95)	11.36 (1.41)	10.12 (.49)	.001
WASI IQ score	118.04 (10.85)	104.12 (8.89)	112.94 (12.12)	<.001
Retention interval (in years)	4.04 (2.05)	4.90 (2.24)	5.00 (2.93)	.314
Intracranial volume (mm <sup>3</sup> ) <sup>d</sup>	1720.03 (146.80)	1688.50 (159.64)	1748.06 (188.30)	.700
Total hippocampus (mm <sup>3</sup> ) <sup>e</sup>	3841.42 (454.46)	3829.81 (471.32)	3560.97 (508.78)	.154
Right hippocampus (mm <sup>3</sup> ) <sup>e</sup>	1961.92 (247.85)	1947.64 (253.77)	1829.90 (267.53)	.245
Right posterior (mm <sup>3</sup> ) <sup>e</sup>	1038.45 (151.52)	1029.05 (119.26)	953.82 (187.19)	.238
Right anterior (mm <sup>3</sup> ) <sup>e</sup>	923.49 (135.60)	918.60 (215.72)	876.08 (165.37)	.684
Left hippocampus (mm <sup>3</sup> ) <sup>e</sup>	1879.50 (223.31)	1882.16 (229.75)	1731.08 (246.97)	.086
Left posterior (mm <sup>3</sup> ) <sup>e</sup>	1022.91 (135.60)	1025.95 (119.75)	965.49 (160.36)	.395
Left anterior (mm <sup>3</sup> ) <sup>e</sup>	856.23 (169.26)	857.03 (231.69)	775.11 (167.55)	.332

<sup>a</sup> 3 CH did not complete MRI scan.

<sup>b</sup> 1 HYPO did not complete MRI scan.

<sup>c</sup> After controlling for appropriate covariates.

<sup>d</sup> Expressed per 1000 mm<sup>3</sup>.

<sup>e</sup> Note: hippocampal volumes reflect group means after controlling for age, sex, and ICV.

subcategory (i.e., event, place, time, perceptual) was the repeated measure, group was the between-subjects variable, and age, sex, and retention interval were covariates. Group differences in accurate visual and non-visual details, and participants' visualization ratings were examined using ANCOVA, controlling for age, sex, and retention interval. Hierarchical linear regressions were used to determine whether participants' visualization ratings predicted total accurate details, accurate visual details or proportion accuracy scores. For these analyses, age, sex, and retention interval were entered as covariates in step 1, dummy-coded group and visualization ratings were entered as predictors in step 2, and dummy-coded group-by-visualization rating interaction variables were entered in step 3.

Group differences in hippocampal volumes were examined using a mixed-factor ANCOVA, with hippocampal sub-region (i.e., right posterior, right anterior, left posterior, left anterior) as the repeated measure, group as the between-subjects variable, and age, sex, and intracranial volume as covariates. In order to assess whether hippocampal volumes better predicted proportion accuracy scores than total episodic details, hierarchical linear regressions were used with age, sex, retention interval, and intracranial volumes entered as covariates in step 1, dummy-coded group variables and hippocampal volumes entered as predictors in step 2, and dummy-coded group-by-hippocampal volume interaction variables entered in step 3. Finally, hierarchical linear regressions were used to examine whether elevated TSH values at diagnosis in CH and maternal TSH levels throughout pregnancy in HYPO predicted proportion accuracy scores after controlling for age, sex, and retention interval. Given the exploratory nature of the present study, an overall significance level of  $p < .05$  was used and a few important trend-level effects (at  $p < .10$ ) are discussed. However, effect sizes are reported and most results approached a significance level of  $p < .01$ .

### 3. Results

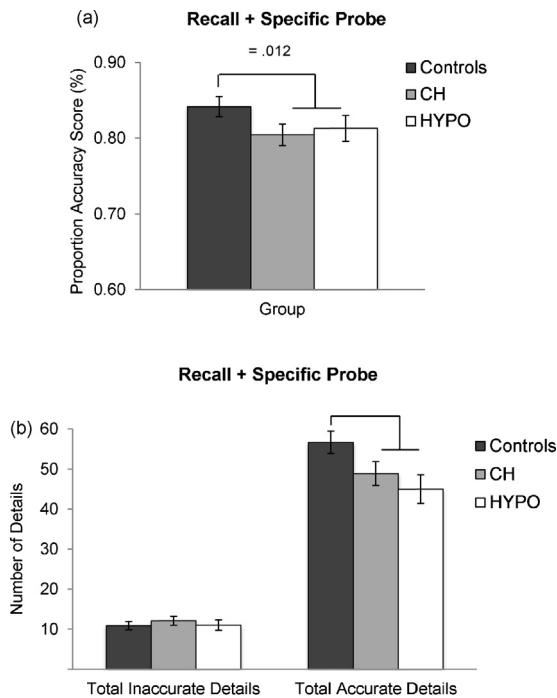
#### 3.1. Demographic measures

Demographic data are presented in Table 3. Three CH participants had learning disabilities and/or ADHD and one HYPO had ADHD. No significant group differences were found in sex distribution or SES. However, the CH group had a higher mean age,  $F(2,71) = 9.73$ ;  $p < .001$ ;  $\eta_p^2 = .22$ , and lower IQ scores,  $F(2,65) = 11.33$ ;  $p < .001$ ;  $\eta_p^2 = .26$ , than HYPO and controls.

#### 3.2. Group differences in AM recall of a staged event

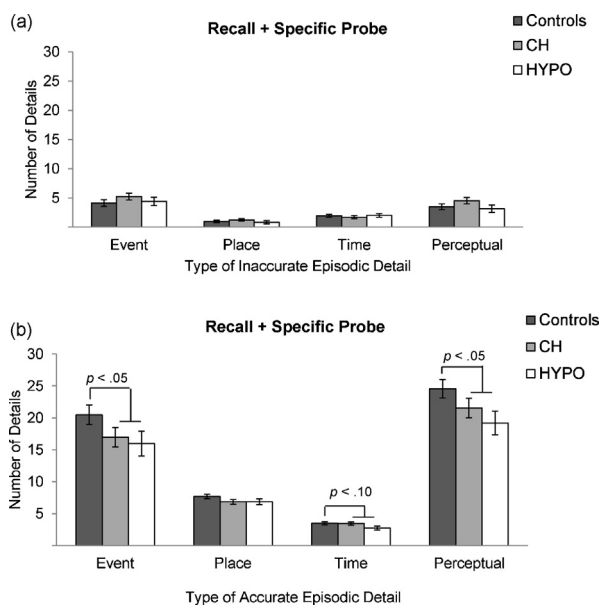
Preliminary analyses revealed that CH and HYPO groups did not differ significantly from each other on any of the AM variables in the present study ( $ps > .05$ ). Thus, to gain statistical power for group comparisons given our small sample size, both groups were collapsed into a single TH-deficient group for all analyses of AM. However, individual group means are presented in all figures of the AM data (e.g., Figs. 2 and 3).

First, group differences in proportion accuracy scores were examined using ANCOVA. A significant main effect of group was found,  $F(1,63) = 6.72$ ;  $p = .012$ ;  $\eta_p^2 = .10$ , indicating that controls had more accurate recollections of the staged event than the combined TH-deficient group (see Fig. 2a for group means). Next, group differences in total inaccurate and total accurate details were examined. Results from the mixed-factor ANCOVA revealed a trend-level main effect of group,  $F(1,63) = 3.51$ ;  $p = .066$ ;  $\eta_p^2 = .05$ , and a significant group by detail category interaction,  $F(1,63) = 10.72$ ;  $p = .002$ ;  $\eta_p^2 = .15$ . Post hoc ANCOVAs indicated that controls recalled significantly more accurate details than the TH-deficient group,  $F(1,63) = 6.90$ ;  $p = .011$ ;  $\eta_p^2 = .10$ ; however, groups did not differ in the number of inaccurate details recalled ( $p = .308$ ; see Fig. 2b for group means).



**Fig. 2.** Group means for proportion accuracy scores (a) and the number of total inaccurate and total accurate details recalled (b).

Group differences in inaccurate episodic detail subcategories (i.e., event, place, time, and perceptual details) using a mixed-factor ANCOVA revealed no significant differences between groups ( $p > .300$ ; see Fig. 3a for group means). However, when groups were similarly compared for differences in *accurate* episodic detail



**Fig. 3.** Group means for the number of details recalled from each inaccurate episodic detail subcategory (a) and each accurate episodic detail subcategory (b).

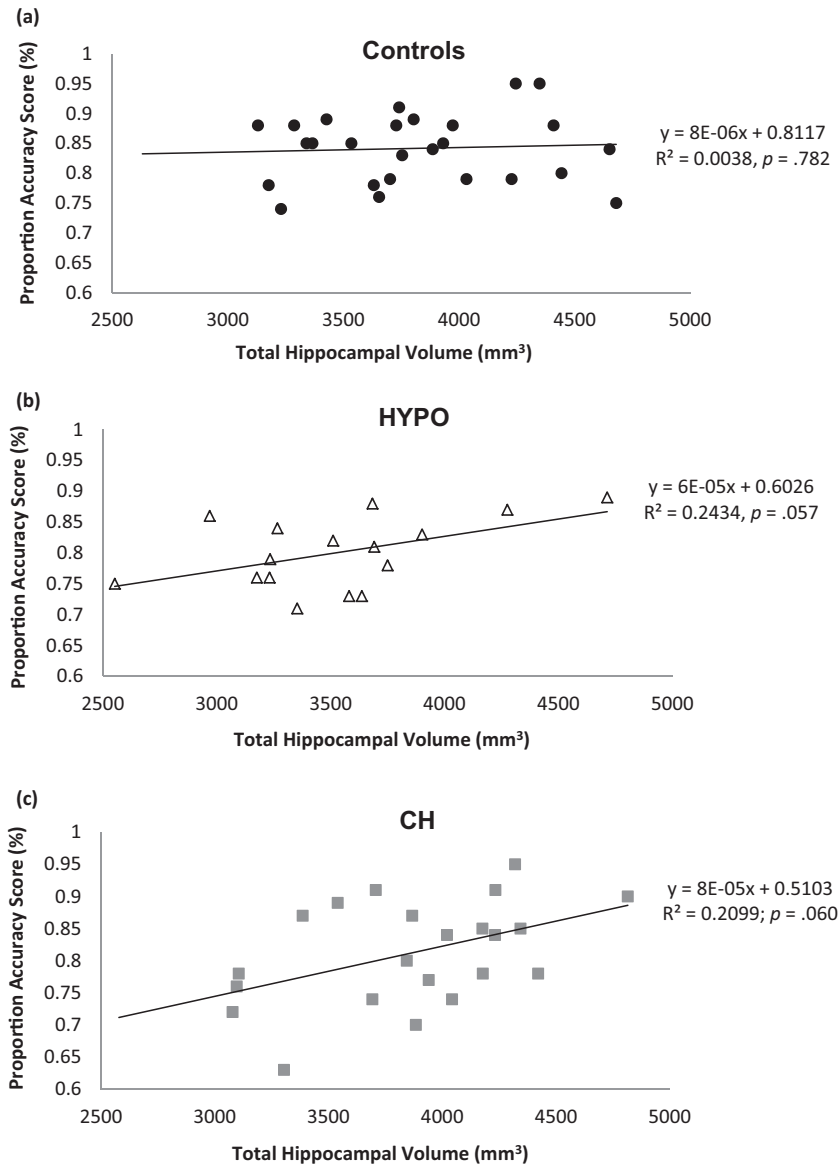
subcategories (i.e., event, place, time, and perceptual details), results indicated a significant main effect of group,  $F(1,63)=7.55$ ;  $p=.008$ ;  $\eta_p^2=.11$ , and a trend-level interaction between group and detail category,  $F(1,63)=2.40$ ;  $p=.093$ ;  $\eta_p^2=.04$ . Post hoc ANCOVAs indicated that controls recalled significantly more accurate event details,  $F(1,63)=4.05$ ;  $p=.049$ ;  $\eta_p^2=.06$ , accurate perceptual details,  $F(1,63)=4.67$ ;  $p=.035$ ;  $\eta_p^2=.07$ , and at a trend-level, accurate place details,  $F(1,63)=3.49$ ;  $p=.066$ ;  $\eta_p^2=.05$ , than the TH-deficient group, but groups did not differ in the number of accurate time details recalled ( $p=.287$ ; see Fig. 3b for group means). Importantly, when the accurate perceptual detail subcategory was further subdivided into accurate visual and accurate non-visual details, a significant group difference was found for accurate visual details,  $F(1,63)=4.77$ ,  $p=.033$ ;  $\eta_p^2=.07$ , but not for non-visual details ( $p=.478$ ). This finding indicates that controls recalled significantly more accurate visual details from the staged event than the TH-deficient group.

### 3.3. Group differences in visualization ratings and its effects on episodic AM

No significant group differences were found in participants' self-report ratings of their ability to visualize the staged event at the time of recollection ( $p=.367$ ). Next, hierarchical linear regressions were used to determine whether participants' visualization ratings predicted total accurate details, accurate visual details, and proportion accuracy scores. Results indicated a significant main effect of participants' visualization ratings only for total accurate details,  $\Delta F(3,61)=3.99$ ;  $\Delta R^2=.13$ ;  $B=211$ ;  $p=.048$ , and accurate visual details,  $\Delta F(3,61)=4.84$ ;  $\Delta R^2=.17$ ;  $B=288$ ;  $p=.010$ . Across all participants, a greater ability to visualize the staged event at the time of recollection significantly predicted a greater number of total accurate details and accurate visual details recalled from the staged event.

### 3.4. Group differences in hippocampal volumes and its effects on episodic AM

Group means and standard deviations for intracranial and hippocampal volumes are presented in Table 3. No significant group differences were found in intracranial volume ( $p=.700$ ). As shown in Table 3, the HYPO group had smaller hippocampal volumes than controls and CH; however, these group differences were not significant using a mixed-factor ANCOVA (e.g.,  $p=.154$  for total hippocampal volumes). Hierarchical linear regressions were used to examine whether total hippocampal volumes better predicted proportion accuracy scores than total episodic details recalled. No significant effect of total hippocampal volume was found for total episodic details. For proportion accuracy scores, however, a significant main effect of total hippocampal volume was found,  $\Delta F(3,56)=3.63$ ;  $\Delta R^2=.14$ ;  $B=313$ ;  $p=.023$ , indicating that smaller hippocampal volumes predicted lower proportion accuracy scores in all groups. As shown in Fig. 4, which displays scatterplots of total hippocampal volumes by proportion accuracy scores for each group, the effect appears to be driven by the CH and HYPO groups.



**Fig. 4.** Scatterplots of total hippocampal volumes by proportion accuracy scores for controls.

Given that total hippocampal volumes were found to significantly predict proportion accuracy scores, additional hierarchical linear regression analyses were conducted to examine whether particular hippocampal subregions (i.e., right posterior, right anterior, left posterior, left anterior) predicted proportion accuracy scores. Results indicated a significant main effect of left anterior hippocampal volumes on proportion accuracy scores,  $\Delta F(3,56) = 3.41$ ;  $\Delta R^2 = .14$ ;  $B = 301$ ;  $p = .031$ , signifying that across all groups, participants with smaller left anterior hippocampal volumes had lower proportion accuracy scores. In addition, a significant interaction between right anterior hippocampal volumes and the dummy-coded group variables was found,  $\Delta F(2,54) = 2.54$ ;  $\Delta R^2 = .09$ ;  $p = .031$  ( $B = 346$ ;  $p = .014$  for the HYPO vs. CON dummy variable; and  $B = 281$ ;  $p = .047$  for the CH vs. CON dummy variable). These significant interactions

indicate that in CH and HYPO groups only, and not controls, smaller right anterior hippocampal volumes significantly predicted lower proportion accuracy scores.

### 3.5. Effects of TH deficiency severity on proportion accuracy scores

Hierarchical linear regressions revealed that higher maternal TSH levels (signifying more severe TH deficiency) during the third trimester significantly predicted lower proportion accuracy scores in the HYPO group,  $\Delta F(1,9) = 5.69$ ;  $\Delta R^2 = .33$ ;  $B = -.479$ ,  $p = .041$ . Maternal TSH levels during the first trimester,  $\Delta F(1,8) = .02$ ;  $\Delta R^2 = .00$ ;  $B = -.032$ ,  $p = .890$ , and the second trimester,  $\Delta F(1,9) = .00$ ;  $\Delta R^2 = .00$ ;  $B = .005$ ,  $p = .985$ , did not significantly predict proportion accuracy scores. When using newborn TSH



levels (i.e., in 18 CH and 11 HYPO), higher TSH levels at birth predicted lower proportion accuracy scores at a trend-level,  $\Delta F(1,24) = 3.60$ ;  $\Delta R^2 = .10$ ;  $B = -.353$ ,  $p = .070$ . However, when using TSH levels at diagnosis only in the CH group ( $N = 23$ ), higher TSH levels at diagnosis did not significantly predict lower proportion accuracy scores,  $\Delta F(1,18) = .33$ ;  $\Delta R^2 = .01$ ;  $B = .141$ ,  $p = .573$ .

#### 4. Discussion

The present study represents the first investigation of AM accuracy in children with early TH deficiency using a staged autobiographical event. Results from the CAI showed that both CH and HYPO groups recalled significantly fewer accurate details and had proportionally less accurate recollections of the staged event than did typically developing controls. Although no previous study has specifically examined AM accuracy in individuals with hippocampal abnormalities, our findings are consistent with other studies investigating AM accuracy in other clinical pediatric populations, such as children with autism or intellectual disability, who report fewer episodic details and provide proportionally less accurate recollections of a staged event than controls (Agnew and Powell, 2004; Bruck et al., 2007). Interestingly, we also found that more severe TH deficiency in the third trimester was associated with lower proportion accuracy scores in the HYPO group. Thus, our results suggest that late prenatal TH deficiency may have long-term consequences for the ability to accurately recall episodic details from past events.

Similar to the results from our recent study showing that CH and HYPO participants have a specific weakness in recalling event and perceptual details from past naturally-occurring autobiographical events (Willoughby et al., 2013a,b), CH and HYPO participants in the present study recalled significantly less accurate event and perceptual details, particularly accurate visual details, about the staged event than controls. These finding suggests that early TH deficiency may lead to a general weakness in the ability to recall event and perceptual details from past events.

Although our results are remarkably similar to the findings from our previous study (Willoughby et al., 2013a,b), it is important to note several important differences when comparing staged versus naturally occurring autobiographical events. For example: (a) the staged event in the present study was a more recent event (i.e., within 4 months) than the naturally occurring AMs recalled in Willoughby et al. (2013a,b), which had a mean retention interval of one year. This difference may have made the staged event easier to recall; (b) participants in the present study were given specific instructions to pay close attention to event and perceptual details during the staged event, which may have led to better encoding and smaller group differences within the event and perceptual subcategories of episodic details; (c) additional probing questions about the staged event were included in the present study, which may have improved recall due to the provision of additional retrieval support, (d) all participants returned to the same location as the staged event for the CAI, which may have provided them with additional retrieval cues, and (e) the

staged event may have been more salient and may have involved more activities, references to time and place, and objects than naturally-occurring events, allowing for more episodic details to be recalled. Despite these differences, significant group effects were still evident in the number of accurate details recalled and proportion accuracy scores in the present study, indicating that early TH deficiency is associated with impaired episodic AM accuracy.

We found no significant group differences in participants' self-report visualization ratings despite the fact that both CH and HYPO groups recalled fewer accurate visual details than controls. However, we also found that across all participants, greater visualization of the staged event at the time of recollection predicted a greater number of total accurate details and accurate visual details recalled. This finding suggests that visual imagery is an important component of accurate episodic AM retrieval.

Although HYPO participants did not exhibit significantly reduced hippocampal volumes in the present study, they did exhibit smaller volumes than CH and controls. The lack of significant group differences in hippocampal volumes in the present study may be due to our small sample size and the fact that treated TH deficiency appears to have only mild effects on hippocampal volume that may be difficult to reliably detect using low resolution scanning (i.e., with 1.5 T scanner). Importantly, our results indicate that smaller hippocampal volumes, particularly smaller anterior hippocampal volumes, significantly predict lower proportion accuracy scores, in the CH and HYPO groups. Previous research indicates three important features of the anterior hippocampus that are relevant to our findings: (a) it may play an essential role in accurate episodic encoding and retrieval (Kircher et al., 2008), (b) it projects to the prefrontal cortex, which also plays a role in AM accuracy by verifying retrieved information (Barbas and Blatt, 1995; Kircher et al., 2008), and (c) it shows less postnatal growth than the posterior hippocampus (Gogtay et al., 2006) and therefore, may be more vulnerable to TH deficiency during gestation. Thus, our findings suggest that reductions in anterior hippocampal volumes associated with late prenatal TH deficiency may be associated with significant weaknesses in episodic AM accuracy. In addition, because hippocampal volumes significantly predicted proportion accuracy scores but not total episodic details across all groups in the present study, our results also suggest that proportion accuracy scores, which assess only accurate AM retrieval and control for total verbal output, may be a particularly sensitive measure of episodic AM impairment and hippocampal dysfunction in patients with hippocampal abnormalities.

Although the present study provides critical new insight into the effects of early TH deficiency on episodic AM accuracy, it is limited for several reasons. First, early TH deficiency impairs fundamental neurobiological processes within other brain regions that are part of the AM neural network, such as the prefrontal cortex, thalamus, and cerebellum (Bernal and Nunez, 1995; Chan and Rovet, 2003). Although the hippocampus plays a critical role in episodic AM and is particularly vulnerable to TH deficiency, damage to these other regions may have also contributed to the weaknesses in AM accuracy observed in children with

early TH deficiency. Second, due to difficulties in recruiting participants, our sample size was small, our groups were not perfectly age-matched, and there were several missing TSH values within the CH and HYPO groups which may have reduced our ability to detect more significant relations in our regression analyses. Third, due to the design of our study it is impossible to disentangle whether the weaknesses in episodic AM accuracy observed in children with early TH deficiency reflect poor encoding of details at the time of the staged event, poor retrieval of these details during the CAI, or a combination of these two impairments. Fourth, we were unable to confirm the euthyroid status of our control mothers given that no routine prenatal screening program for maternal hypothyroidism during pregnancy was in existence in Ontario at the time the mothers of controls were pregnant. Thus, our results should be interpreted with these limitations in mind and further investigations are required to replicate our findings.

## 5. Conclusion

The present study provides the first indication that children with early-treated CH and children born to women with treated maternal hypothyroidism exhibit difficulties in recalling accurate episodic details from past events. This weakness appears to be associated with both reduced anterior hippocampal volumes and more severe TH deficiency during late gestation. In addition, hippocampal volumes were found to significantly predict proportion accuracy scores, but not the total number of episodic details recalled, suggesting that proportion accuracy scores may provide a more sensitive measure of children's episodic AM ability, as well as hippocampal dysfunction, in clinical pediatric populations. However, given that hippocampal volumes provide no insight into the functional integrity of the hippocampus, additional investigations using functional MRI are required in order to examine whether early TH deficiency is specifically associated with abnormal hippocampal function during episodic AM retrieval. Overall, this study provides critical new insight into the effects of early TH deficiency on AM accuracy and the hippocampus. In addition, it has important implications for the treatment and management of maternal and congenital hypothyroidism given that deficits in AM accuracy were found despite the fact that all cases received some degree of TH treatment.

## Conflict of interest statement

All authors declare that there is no conflict of interest.

## Acknowledgements

This research was supported by a grant from the Canadian Institutes of Health Research (MOP 49488) to JR and MPM; KW was supported through an Ontario Graduate Scholarship and a Hospital for Sick Children RESTRACOMP scholarship. We wish to thank Dr. Benjamin Zendel for his assistance in the conceptualization of the staged event; Anishka Leis, Victoria Martin, and Rosie Bell for recruiting participants and assistance with conducting the study;

Jovanka Skocic for assistance in performing the tracings for reliability; Susan Blaser for neuroradiological evaluations; Tammy Rayner, Garry Detzler, and Ruth Weiss for scanning. Finally, we want to thank the children and their parents for their continued participation and involvement in this research.

## References

- Addis, D.R., Moscovitch, M., McAndrews, M.P., 2007. Consequences of hippocampal damage after the autobiographical memory network in left temporal lobe epilepsy. *Brain* 130, 2327–2342.
- Agnew, S.E., Powell, M.B., 2004. The effect of intellectual disability on children's recall of an event across different question types. *Law and Human Behavior* 28, 273–294.
- Barbas, H., Blatt, G.J., 1995. Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus* 5, 511–533.
- Bernal, J., Nunez, J., 1995. Thyroid hormones and brain development. *European Journal of Endocrinology* 133, 390–398.
- Brassen, S., Weber-Fahr, W., Sommer, T., Lehmsbeck, J.T., Braus, D.F., 2006. Hippocampal-prefrontal encoding activation predicts whether words can be successfully recalled or only recognized. *Behavioural Brain Research* 171, 271–278.
- Bruck, M., London, K., Landa, R., Goodman, J., 2007. Autobiographical memory and suggestibility in children with autism spectrum disorder. *Development and Psychopathology* 19, 73–95.
- Cabeza, R., Prince, S.E., Daselaar, S.M., Greenberg, D.L., Budde, M., Dolcos, F., et al., 2004. Brain activity during episodic retrieval of autobiographical and laboratory events: an fMRI study using a novel photo paradigm. *Journal of Cognitive Neuroscience* 16, 1583–1594.
- Chan, S., Rovet, J., 2003. Thyroid hormones in fetal central nervous system development. *Fetal & Maternal Medicine Review* 13, 177–208.
- Eldridge, L.L., Knowlton, B.J., Furmanski, C.S., Bookheimer, S.Y., Engel, S.A., 2000. Remembering episodes: a selective role for the hippocampus during retrieval. *Nature Neuroscience* 3, 1149–1152.
- Gilbert, M.E., Paczowski, C., 2003. Propylthiouracil (PTU)-induced hypothyroidism in the developing rat impairs synaptic transmission and plasticity in the dentate gyrus of the adult hippocampus. *Developmental Brain Research* 145, 19–29.
- Glinoe, D., Spencer, C.A., 2010. Serum TSH determinations in pregnancy: how, when, and why? *Nature Reviews Endocrinology* 6, 526–529.
- Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., et al., 2006. Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America* 101, 8174–8179.
- Hollingshead, A.B., 1975. Four-Factor Index of Social Status. Yale University, New Haven, CT (unpublished manuscript).
- Hyman, I.E., Loftus, E.F., 1998. Errors in autobiographical memory. *Clinical Psychology Review* 18, 933–947.
- Kircher, T., Weis, S., Leube, D., Freymann, K., Erb, M., Jessen, F., 2008. Anterior hippocampus orchestrates successful encoding and retrieval of non-relational memory: An event-related fMRI study. *European Archives of Psychiatry and Clinical Neuroscience* 258, 363–372.
- Koriat, A., Goldsmith, M., 1994. Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General* 123, 297–315.
- Lavado-Autric, R., Auso, E., Garcia-Velasco, J.V., del Carmen Arufe, M., Escobar del Rey, F., Berbel, P., Morreale de Escobar, G., 2003. Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny. *Journal of Clinical Investigation* 111, 1073–1082.
- Levine, B., Svoboda, E., Hay, J.F., Winocur, G., Moscovitch, M., 2002. Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychology and Aging* 17, 677–689.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1, 30–46.
- Montaldi, D., Mayes, A.R., Barnes, A., Hadley, D.M., Patterson, J., Wyper, D.J., 2001. Effects of level of retrieval success on recall-related frontal and medial temporal lobe activations. *Behavioural Neurology* 13, 123–131.
- Moscovitch, M., 2008. The hippocampus as a "stupid" domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology* 62, 62–79.

- Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three dimensional analysis software: minimizing the discrepancies between laboratories. *Cerebral Cortex* 10, 433–442.
- Roebers, C.M., Schneider, W., 2000. The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology* 14, 509–526.
- Skinner, E.I., Fernandes, M.A., Grady, C.L., 2009. Memory networks supporting retrieval effort and retrieval success under conditions of full and divided attention. *Experimental Psychology* 56, 386–396.
- Sperling, R., Chua, E., Cocchiarella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D.L., Albert, M., 2003. Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage* 20, 1400–1410.
- Stark, C.E., Squire, L.R., 2000. Functional magnetic resonance imaging (fMRI) activity in the hippocampal region during recognition memory. *Journal of Neuroscience* 20, 7776–7781.
- St-Laurent, M., Moscovitch, M., Levine, B., McAndrews, M.P., 2009. Determinants of autobiographical memory in patients with unilateral temporal lobe epilepsy or excisions. *Neuropsychologia* 47, 2211–2221.
- Svoboda, E., McKinnon, M.C., Levine, B., 2006. The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia* 44, 2189–2208.
- Wechsler, D., 1999. *Wechsler Abbreviated Scale of Intelligence*. Psychological Corporation, New York.
- Weiss, A.P., Dewitt, I., Goff, D., Ditman, T., Heckers, S., 2005. Anterior and posterior hippocampal volumes in schizophrenia. *Schizophrenia Research* 73, 103–112.
- Wheeler, S.M., Willoughby, K.A., McAndrews, M.P., Rovet, J.F., 2011. Hippocampal size and memory functioning in children and adolescents with congenital hypothyroidism. *Journal of Clinical Endocrinology & Metabolism* 96, E1427–E1434.
- Willoughby, K.A., (Doctoral dissertation) 2011. *Effects of Early Thyroid Hormone Deficiency on Autobiographical Memory and Hippocampal Structure and Function During Late Childhood and Early Adolescence*, Retrieved from ProQuest Dissertations and Theses, Publication number NR78051.
- Willoughby, K.A., Desrocher, M., Levine, B., Rovet, J.F., 2012. Episodic and semantic autobiographical memory and everyday memory during late childhood and early adolescence. *Frontiers in Developmental Psychology* 3, 1–15.
- Willoughby, K.A., McAndrews, M.P., Rovet, J., 2013a. Effects of early thyroid hormone deficiency on children's autobiographical memory performance. *Journal of the International Neuropsychological Society* 19, 419–429.
- Willoughby, K.A., McAndrews, M.P., Rovet, J., 2013b. Effects of maternal hypothyroidism on offspring hippocampus and memory. *Thyroid* (Epub ahead of print).
- Zoeller, R.T., Rovet, J., 2004. Timing of thyroid hormone action in the developing brain: clinical observations and experimental findings. *Journal of Neuroendocrinology* 16 (10), 809–818.