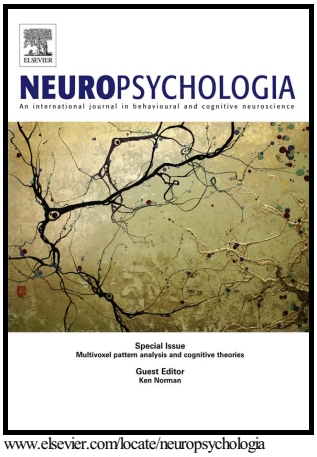
Author's Accepted Manuscript

Distinguishable memory retrieval networks for collaboratively and non-collaboratively learned information

Flora Vanlangendonck, Atsuko Takashima, Roel M. Willems, Peter Hagoort



www.elsevier.com/locate/neuropsychologia

PII: S0028-3932(17)30473-6
DOI: https://doi.org/10.1016/j.neuropsychologia.2017.12.008
Reference: NSY6599

To appear in: *Neuropsychologia*

Received date: 30 July 2017
Revised date: 19 November 2017
Accepted date: 3 December 2017

Cite this article as: Flora Vanlangendonck, Atsuko Takashima, Roel M. Willems and Peter Hagoort, Distinguishable memory retrieval networks for collaboratively and non-collaboratively learned information, *Neuropsychologia,* https://doi.org/10.1016/j.neuropsychologia.2017.12.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Distinguishable memory retrieval networks for collaboratively and non-collaboratively learned information

Flora Vanlangendonck[1,2], Atsuko Takashima[1,3], Roel M. Willems[1,3,4], & Peter Hagoort[1,3]

1 Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands,

2 International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands,

3 Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands,

4 Centre for Language Studies, Radboud University Nijmegen, The Netherlands

Corresponding author. Flora Vanlangendonck, Donders Institute for Brain Cognition and Behaviour, Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, 6500 HB Nijmegen, f.vanlangendonck@donders.ru.nl

Accepted manuscript

ACCEPTED MANUSCRIPT

Abstract

Learning often occurs in communicative and collaborative settings, yet almost all research into the neural basis of memory relies on participants encoding and retrieving information on their own. We investigated whether learning linguistic labels in a collaborative context at least partly relies on cognitively and neurally distinct representations, as compared to learning in an individual context. Healthy human participants learned labels for sets of abstract shapes in three different tasks. They came up with labels with another person in a collaborative communication task (*collaborative condition*), by themselves (*individual condition*), or were given pre-determined unrelated labels to learn by themselves (*arbitrary condition*). Immediately after learning, participants retrieved and produced the labels aloud during a communicative task in the MRI scanner. The fMRI results show that the retrieval of collaboratively generated labels as compared to individually learned labels engages brain regions involved in understanding others (*mentalizing* or *theory of mind*) and autobiographical memory, including the medial prefrontal cortex, the right temporoparietal junction and the precuneus. This study is the first to show that collaboration during encoding affects the neural networks involved in retrieval.

Accepted manuscript

ACCEPTED MANUSCRIPT

Keywords: fMRI, communication, social interaction, hippocampus, language, memory, collaboration

Accepted manuscript

Introduction

Learning often occurs in communicative and collaborative settings, yet almost all research into the neural basis of memory relies on participants encoding and retrieving information on their own. Intuitively, it may seem only a small step to extrapolate that the memory systems that support encoding and retrieval in an individual context are also involved in encoding and retrieving in a collaborative context. However, a study with hippocampal amnesia patients suggests that learning in a collaborative context and learning in an individual context may at least partly rely on cognitively and neurally distinct processes.

Duff, Hengst, Tranel, & Cohen (2006) instructed patients with hippocampal amnesia to perform multiple rounds of the "tangram task", a classical referential communication task (Clark & Wilkes-Gibbs, 1986). In this task, a director and a matcher each have a set of abstract figures, but their view of the other person's figures is occluded by a barrier. The director instructs the matcher to rearrange their figures to match the director's order. Over multiple rounds of this task, healthy participants converge on a set of labels that are used to refer to the abstract figures. These labels become shorter and more efficient with each repetition of the task (Clark & Wilkes-Gibbs, 1986). Interestingly, the amnesic patients could successfully learn novel labels for the figures over multiple rounds of the communicative tangram task,

ACCEPTED MANUSCRIPT

but they were unable to learn arbitrary relations between labels and pictures in an individual paired-associate learning task. Duff et al. proposed that the successful learning in the communicative task occurred through tuning of existing conceptual, semantic and visual networks and did not require hippocampus-dependent learning of new associations. Their findings thus suggest that learning labels in a collaborative context may rely less on hippocampally oriented memory systems, as compared to learning such labels individually.

In addition to this evidence from lesion work, behavioral research also suggests that collaboration affects memory processes (Rajaram & Pereira-Pasarin, 2010). Both collaborative encoding and retrieval have been found to impair memory recall compared to individual learning (Barber, Rajaram, & Aron, 2010; Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997). Furthermore, during collaborative retrieval of individually encoded memories, memories can be reinforced as a result of repetition (Blumen & Rajaram, 2008; Rajaram & Pereira-Pasarin, 2007; Weldon & Bellinger, 1997), forgotten if they are not mentioned during the interaction (Coman, Manier, & Hirst, 2009; Cuc, Koppel, & Hirst, 2007), and false memories can be transmitted between people (social memory contagion; Roediger, Meade, & Bergman, 2001). Despite this behavioral literature on collaborative memory, very little is known about the neural mechanisms that support collaborative memory encoding and retrieval.

In the current study, we directly compared the neural representation of collaboratively generated linguistic labels to memory for self-generated or pre-determined labels that were encoded individually in the healthy population. During a behavioral session, participants generated labels for abstract figures together with another person in a collaborative communication task (*collaborative condition*), by themselves (*individual condition*), or were given pre-determined arbitrary labels to learn by themselves (*arbitrary condition*). Immediately after learning, participants completed a communicative task in the MRI scanner during which they retrieved the labels they had learned during the training session. This

Accepted manuscript

ACCEPTED MANUSCRIPT

design allowed us to directly compare both behavioral and neural activation patterns relating to the retrieval of labels learned in the collaborative, individual and arbitrary contexts.

In line with the results from the hippocampal amnesia patients, we expected that the retrieval of pre-determined arbitrary labels should rely strongly on the hippocampus, while the retrieval of the collaboratively generated labels may be less hippocampus-dependent. More specifically, we hypothesized that the retrieval of labels learned in the collaborative condition would partly rely on brain areas involved in *mentalizing* or *theory of mind*, i.e. people's ability to track and understand what other people think or feel. These areas, including the medial prefrontal cortex, the temporoparietal junction and the precuneus, are consistently found in neuroimaging studies of social cognition and interaction (Amodio & Frith, 2006; Frith & Frith, 2006; Frith & Frith, 2010; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). We therefore hypothesized that these areas should be more active in the collaborative condition than in the individual condition. In addition, we expected that retrieving self-generated labels (collaborative and individual conditions) as compared to arbitrary labels may activate areas known to be important for storing semantic associations such as the left angular gyrus, middle temporal gyrus and temporal poles (Binder, Desai, Graves, & Conant, 2009). Behaviorally, we expected to replicate the results of previous tangram task studies during the behavioral practice phase. That is, participants were expected to produce shorter and more efficient descriptions with each round of this task (Clark & Wilkes-Gibbs, 1986).

Accepted manuscript

Materials and methods

Participants

36 pairs of native Dutch speakers participated in the experiment. Participants signed up individually through the university research participation system. They did not know each other before the start of

ACCEPTED MANUSCRIPT

the experiment. All participants had normal or corrected-to-normal vision and no history of neurological disease. They gave written informed consent before the start of the experiment. For each pair, one participant was assigned the role of director and the other the role of matcher. Data from one pair were excluded due to excessive movement by the participant in the MRI scanner. The results of the remaining seventy participants (directors: 7 men, 21.43 years old, range 18-33 years; matchers: 6 men, 21.34 years old, range 18-31 years) are reported below.

Procedure

The experiment lasted approximately two hours and consisted of two parts: a behavioral session and an fMRI session. The fMRI session took place immediately after the behavioral session. Before the start of the experiment, participants were assigned to the roles of director or matcher for the duration of the experiment by means of a coin toss. During the behavioral session, participants were seated behind different monitors in the same room. The monitors were separated by a screen, so the participants could not see each other or each other's monitor. The director learned sixty labels for pictures in three different tasks: a collaborative task, an individual task and an arbitrary task. The matcher studied the same picture labels as the director during the behavioral session. Participants received written instructions before each task. To account for possible task order effects on learning and retrieval, or the specifics of the pictures, the order of the tasks during the behavioral session as well as the assignment of the picture sets to the three tasks were counterbalanced. During the fMRI session, the director performed a referential communication task with the matcher. We recorded the directors' vocal responses during both sessions. An overview of the tasks is depicted in Figure 1.

Accepted manuscript

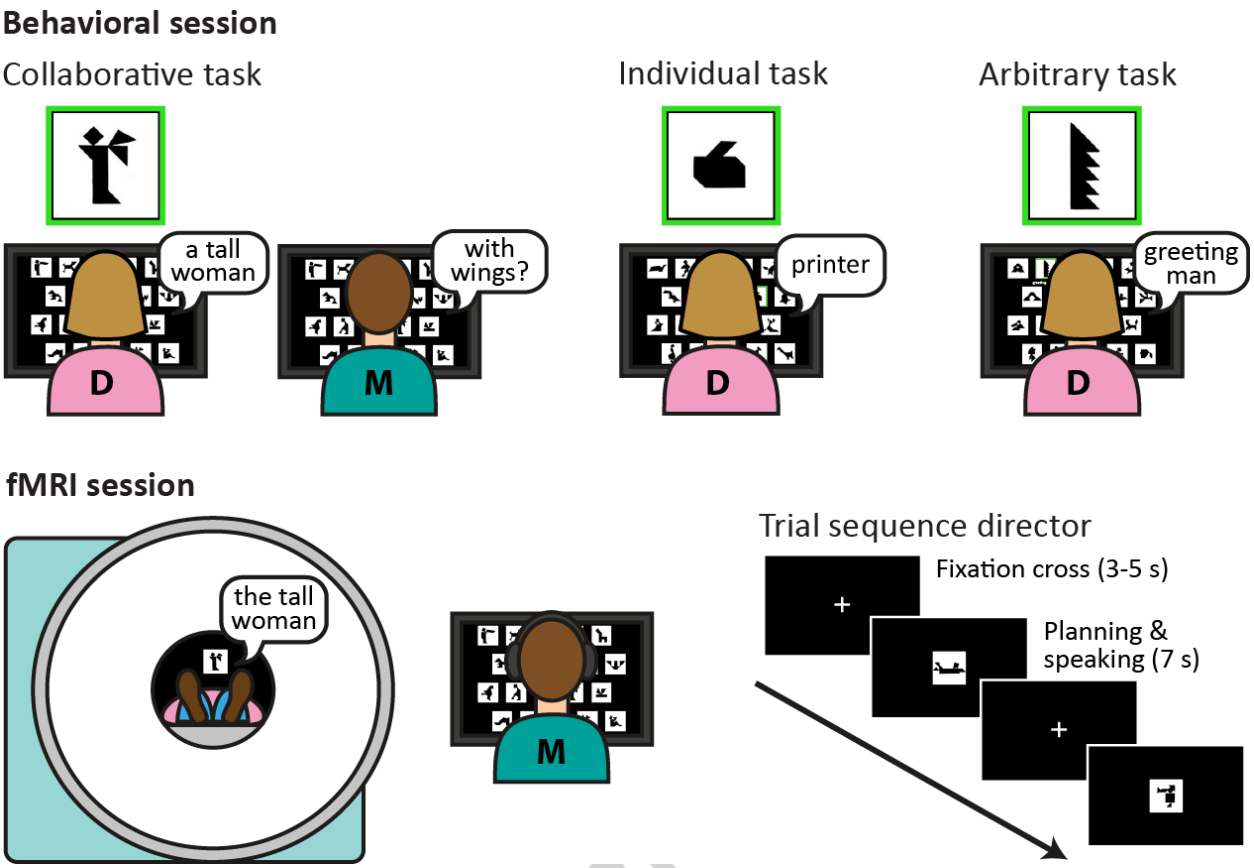


Figure 1: Set-up of the experiment. The top row depicts the three tasks participants performed during the behavioral learning session. The matcher only participated in the collaborative task, but learned the same sets of labels as the director. The described tangram pictures are shown at the top to illustrate the relationship between the pictures and the labels. After completing these three tasks, the director retrieved all previously learned labels in the MRI scanner, while the matcher tried to click on the described pictures. The trial sequence depicts the task in the MRI scanner from the director's point of view. Each picture was presented for seven seconds, during which the director planned and produced his/her response out loud. The director was instructed to press a button just before speaking.

During the collaborative task, the director and matcher came up with labels together in a collaborative and communicative setting. They were seated at different computer monitors. Each participant saw the same set of twenty pictures on their monitor. On each trial, the director was cued to describe one of the

Accepted m

ACCEPTED MANUSCRIPT

pictures, and the matcher clicked on the picture that he/she thought was described by the director. We did not instruct the directors on how to phrase their descriptions (e.g., description length or specific strategies), but explained that they could communicate freely with the matcher during the entire task, and could discuss each picture as long as needed until the matcher had selected the correct picture. Both participants received feedback on screen ("correct" or "wrong") when the matcher clicked on a picture. Once the matcher had clicked on the described picture, the director was cued to describe the next picture. Participants completed four rounds of the collaborative task. In each round of the task, the director was cued to describe each of the twenty pictures once. The locations of the pictures on screen as well as the order in which they were cued were randomized before each round. Afterwards, the director received a list containing all pictures and wrote down the labels they had come up with below the corresponding pictures.

During the individual task, the director came up with labels for pictures by him- or herself. The matcher did not participate in the individual task, but was seated in the same room. In this task, the director saw a different set of twenty pictures on their monitor. The director was instructed to come up with a label for each picture that he/she could easily remember. He/she moved from picture to picture by pressing a button. In each round of the task, the director was cued to describe each of the twenty pictures once out loud. Directors completed four rounds of this task. The locations of the pictures on screen as well as the order in which they were cued were randomized before each round. Afterwards, the director received a list containing all pictures and wrote down the labels he/she had come up with below the corresponding pictures.

During the arbitrary task, the director studied a set of predetermined, arbitrary picture labels. The director was again presented with a different set of twenty pictures on his/her monitor. He/shewas instructed to learn a predetermined label for each picture. The matcher did not participate in this task.

Accepted manuscript

ACCEPTED MANUSCRIPT

The director could move from picture to picture by pressing a button. When he/she pressed the button once, a cue appeared around one of the pictures. When they pressed the button again, the label they had to learn appeared below the picture. In the first round of the task, they could look at the pictures and labels without describing the pictures. From the second round onwards, they were instructed to say the label they had learned for each picture out loud when the cue appeared. When they were sure they knew the label associated with a specific picture, they could remove the picture from the practice set by pressing a second button. The directors practiced the labels until they had removed all labels from the practice set. Then they were asked to label all pictures once more to check their performance.

The matcher had to study all the labels the director had learned during the behavioral session before the start of the fMRI session to allow participants to successfully perform the referential communication game together during the fMRI session. The matcher therefore learned the same sets of labels as the director during the behavioral session. In the collaborative task, the director and matcher established and learned the labels together over the course of the task. The matcher received the labels the director learned in the other tasks on paper and studied these in silence. He/she was given the list of pictures and corresponding labels the director wrote down after the individual task and a list of the pictures and corresponding arbitrary labels the director had studied. Both participants were informed that they had learned the same labels during the behavioral session before the start of the fMRI session. This allowed the director to use the labels he/she had learned during the behavioral session to successfully instruct the matcher during the task in the MRI scanner.

During the fMRI session, the director lay in the MRI scanner, while the matcher was seated at a computer outside the scanner room. The director spoke through a noise-cancelling microphone and could press a button on a button box; the matcher wore headphones and used a mouse to click on pictures. Participants were told that they would play a communication game in the scanner in which the

Accepted manuscript

ACCEPTED MANUSCRIPT

matcher had to select the pictures named by the director. Before the director went into the scanner, we explained to both participants that they had learned the same sets of labels. On each trial, the director was presented with one of the sixty pictures he/she had studied during the behavioral session. The director was instructed to press a button once he/she remembered the associated label and to then say this label out loud. The button presses provided us with a measure of the director's planning duration per trial. At the same time, the matcher saw 20 pictures, organized per set, and tried to click on the described picture.

Trials for the director and matcher always started simultaneously, but differed in length. The director's trials lasted 7 seconds and were followed by a jittered ITI of 3-5 seconds. The matcher's trials lasted 9.5 seconds and were followed by an ITI of 0.5-3 seconds (depending on the director's ITI) to give them sufficient time to select the correct picture. All pictures were shown twice during the MRI task in randomized order, resulting in 120 trials in total.

Materials

The materials consisted of sixty abstract tangram figures, which were adapted from Read (1965). We ran an online pretest in which 20 participants rated 83 tangram pictures. They were asked to come up with a label for each picture and rated how difficult it was to describe each picture on a scale from 1 to 5. Based on these ratings, we created three sets of twenty tangram pictures each that were matched based on the difficulty ratings and the predominant semantic category of the responses (animal, human or object responses).

The labels directors studied in the arbitrary task were selected from the labels participants produced in the same pretest as well as in a previous pretest in which we used the same task but a different set of tangram pictures. We only included labels that were generated for pictures that were not selected for the final stimulus set. The three sets of arbitrary labels were matched in terms of semantic categories as

Accepted manuscript

ACCEPTED MANUSCRIPT

well as the average number of words per label. Each arbitrary label was then linked to an unrelated tangram picture. An example of an arbitrary picture-label pair is shown in the arbitrary example trial in Figure 1.

Data acquisition and analysis

fMRI data acquisition and preprocessing. Participants were scanned in a Siemens 1.5T Avanto MRI scanner using a 32-channel head coil. Functional images were acquired using a T2*-weighted gradient multi-echo planar imaging sequence (TR = 2320 ms; TE1 at 9.4 ms, TE2 at 21 ms, TE3 at 33 ms, TE4 at 44 ms, TE5 at 56 ms; 37 slices; ascending slice order; 3 mm slice thickness; 0.51 mm slice gap; 64 x 64 matrix size; 212 x 212 mm FOV; 90° flip angle and 3.3 x 3.3 x 3 mm voxel size). In addition, T1-weighted anatomical scans with 1 mm isotropic resolution were acquired (TR = 2250 ms; TE = 2.95 ms; 15° flip angle; 256 x 256 x 176 mm FOV).

We acquired 35 pre-scans before the start of the task in the MRI scanner. These scans were used to calculate the optimal weighting to combine the five echoes to one value per volume for each voxel, and this weighting matrix was applied to the remaining functional scans (Poser, Versluis, Hoogduin, & Norris, 2006). Preprocessing was done in SPM8 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm/). The preprocessing of the functional images consisted of slice timing correction to the onset of the middle slice, coregistration of the functional images to the T1 based on the subject-mean functional image, normalization to MNI space (resulting voxel size 2 x 2 x 2 mm) and spatial smoothing using a 3-dimensional isotropic Gaussian smoothing kernel (full-width half-maximum = 8 mm).

Behavioral data analysis. We were interested in testing whether directors' responses became shorter with each round of the collaborative task during the behavioral practice session. We therefore calculated how long each trial of the collaborative task lasted based on the time between stimulus

Accepted manuscript

ACCEPTED MANUSCRIPT

presentation and the correct click by the matcher. In addition, we transcribed and coded all responses produced by the director during the fMRI task and counted the number of words per response. The responses were coded as correct or incorrect by comparing them to the labels the director had come up with (collaborative and individual conditions) or studied (arbitrary condition) during the behavioral session. The coding was used to calculate the director's accuracy during the fMRI task. Furthermore, we used the button press responses from the director in the scanner to estimate planning durations during the fMRI session. Two participants forgot to press the button during the task in the MRI scanner and one participant pressed the button at the wrong time (after instead of before naming the picture labels). We manually calculated the planning durations for these three participants by inspecting the voice onset time in the audio recordings of their responses using Praat (Broersma, 2001), and used these values for the analyses. For the analysis of the number of words per label and the planning durations, we excluded incorrect trials, trials without a button press and trials with planning durations that were more than 3 SD from the mean per condition and the mean per participant (9.83 % of trials). Greenhouse-Geisser correction was applied when the sphericity assumption was violated and Holm–Bonferroni correction for multiple comparisons was used for all reported post-hoc tests. Finally, although our main focus is on the directors' performance and responses, we calculated the matcher's accuracy during the fMRI session, i.e. whether the matcher clicked on the intended picture on each trial. Clicking responses from one matcher were not recorded due to a technical problem and therefore not included in the analysis.

Accepted manuscript

fMRI analysis. We included five variables in the first-level general linear models: collaborative retrieval, individual retrieval, arbitrary retrieval, speaking and trials of no interest. The condition of no interest regressor included trials without a button press and trials with responses that did not match the labels learned during the behavioral session. Note that the condition labels (collaborative, individual and arbitrary) refer to the tasks during the behavioral session; the fMRI task was always done together by the director and the matcher and the trial presentation was the same for all three conditions. The

ACCEPTED MANUSCRIPT

retrieval regressors were stick functions time-locked to picture onset. We included the planning duration associated with each trial as a linear parametric modulator. The speaking regressor was modeled as a stick function with the onset determined by the button press for each trial. We included the number of words per trial as a linear parametric modulator alongside the speaking regressor. The parametric modulators were included in the GLM to exclude possible confounds due to differences in planning durations or in the number of words used in the responses, which may affect the BOLD response. The collaborative, individual and arbitrary retrieval regressors should therefore only reflect successful memory retrieval processes. Regressors were convolved with the hemodynamic response function. In addition, we included 24 nuisance regressors: the six realignment parameters, their square, their first derivative, and the realignment parameters used to realign the previous volume. Individual t-contrasts of interest were created and used in second-level random-effect analyses. Group analyses were performed using one-sample t-tests. Whole-brain results were corrected for multiple comparisons by combining a $p < 0.001$ voxel-level threshold with a cluster extent threshold determined by means of a Monte Carlo simulation with 2500 iterations, after estimation of the smoothness of the data (Slotnick, Moo, Segal, & Hart, 2003). This revealed that clusters of 41 voxels or larger indicated statistically significant effects at the $p < 0.05$ level, corrected for multiple comparisons. Since we had specific hypothesis about the involvement of the hippocampus, we also ran a region of interest (ROI) analysis restricting the search field to the anatomically defined left and right hippocampus using the MarsBaR AAL ROI of this region (Brett, Anton, Valabregue, & Poline, 2002).

Accepted manuscript

Results

Behavioral results

First, we tested whether the tasks during the behavioral session differed in duration. In addition, we tested whether we could replicate the results of previous studies that used the tangram task. These

ACCEPTED MANUSCRIPT

studies typically find that directors' descriptions become shorter and more efficient over time (e.g., Clark & Wilkes-Gibbs, 1986; Brennan & Clark, 1996). We therefore compared the time per picture across the four rounds of the collaborative training task. Second, we compared the planning durations, accuracy and number of words per label between conditions in the fMRI session.

Behavioral session. The tasks during the behavioral session were all self-paced, and there were differences in duration between the tasks (see Figure 2A). A repeated-measures ANOVA with condition as within-subject factor revealed a significant main effect of condition (collaborative, individual or arbitrary) on the total duration per task, $F(2,68) = 62.99$, $p < 0.001$. Post-hoc comparisons revealed significant differences between all conditions ($p < 0.001$ for collaborative vs. individual and individual vs. arbitrary, $p < 0.05$ for collaborative vs. arbitrary).

To see whether we could replicate previous behavioral studies that used similar collaborative tasks, we compared the duration per picture over subsequent rounds of the collaborative task. The duration per picture decreased with each round of the collaborative task (round 1: 27 s, round 2: 11.10 s, round 3: 6.76 s, round 4: 5.89 s; see Figure 2B). A repeated-measures ANOVA with round as within-subject factor revealed a main effect of round, $F(3,102) = 120.58$, $p < 0.001$. A polynomial trend analysis revealed significant linear and quadratic trends across rounds (linear: $F(1,34) = 145.08$, $p < 0.001$; quadratic: $F(1,34) = 92.85$, $p < 0.001$). Our results thus replicate previous studies that found that directors' descriptions become more efficient and shorter with each round of the tangram task. We also looked at the duration per picture in the individual task and observed a similar pattern in this task: the duration per picture decreased with each round of the task (round 1: 8.79 s, round 2: 3.94 s, round 3: 2.98, round 4: 2.64 s). A repeated-measures ANOVA with round as within-subject factor revealed a main effect of round, $F(3,102) = 86.33$, $p < 0.001$.

Accepted manuscript

ACCEPTED MANUSCRIPT

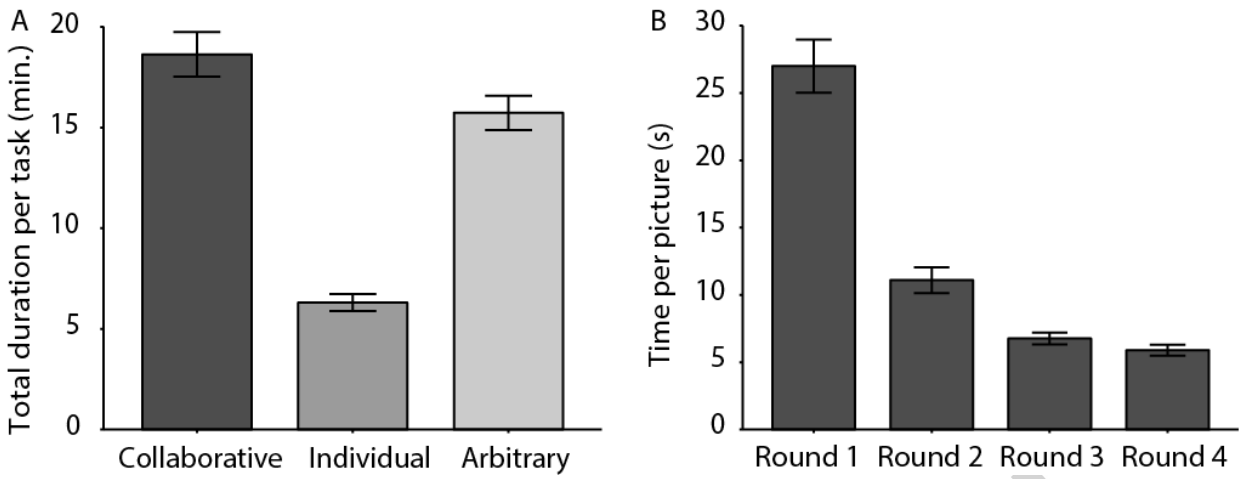


Figure 2: Behavioral results from the behavioral session: A) total time per training task , and B) time per picture per round in the collaborative task. Error bars indicate standard error of the mean.

fMRI session. We performed repeated-measures ANOVAs with condition (collaborative, individual or arbitrary) as within-subject factor and the director's accuracy, planning durations and number of words per label as dependent measures (see Figure 3). We found a significant main effect of condition on accuracy, $F(2,68) = 34.42$, $p < 0.001$. Post-hoc comparisons revealed that accuracy was significantly higher in the collaborative condition compared to the individual condition ($t(34) = 2.81$, $p < 0.01$) and in the individual compared to the arbitrary condition ($t(34) = 4.91$, $p < 0.001$). We found a significant main effect of condition on planning durations in the fMRI task, $F(2, 68) = 72.26$, $p < 0.001$. Post-hoc comparisons revealed significantly shorter planning durations in the individual compared to the arbitrary condition, $t(34) = 9.40$, $p < 0.001$. The difference between the collaborative and the individual conditions was not significant, $t(34) = 1.55$, $p = 0.13$. We found a significant main effect of condition on the number of words per label in the MRI task, $F(2, 68) = 26.08$, $p < 0.001$. Post-hoc comparisons revealed that responses contained significantly more words in the collaborative condition compared to the individual condition ($t(34) = 4.61$, $p < 0.001$) and in the individual condition compared to the arbitrary condition ($t(34) = 2.70$, $p < 0.05$).

Accepted manuscript

ACCEPTED MANUSCRIPT

Matchers' accuracy during the fMRI session was highest for the collaboratively learned labels (collaborative: M = 86.91%, SEM: 1.87%; individual: 78.60%, SEM: 2.61%; arbitrary: 60.15%, SEM: 3.01%). Note that the matchers' performance was highly dependent on the directors' accuracy and planning durations, which differed across conditions.

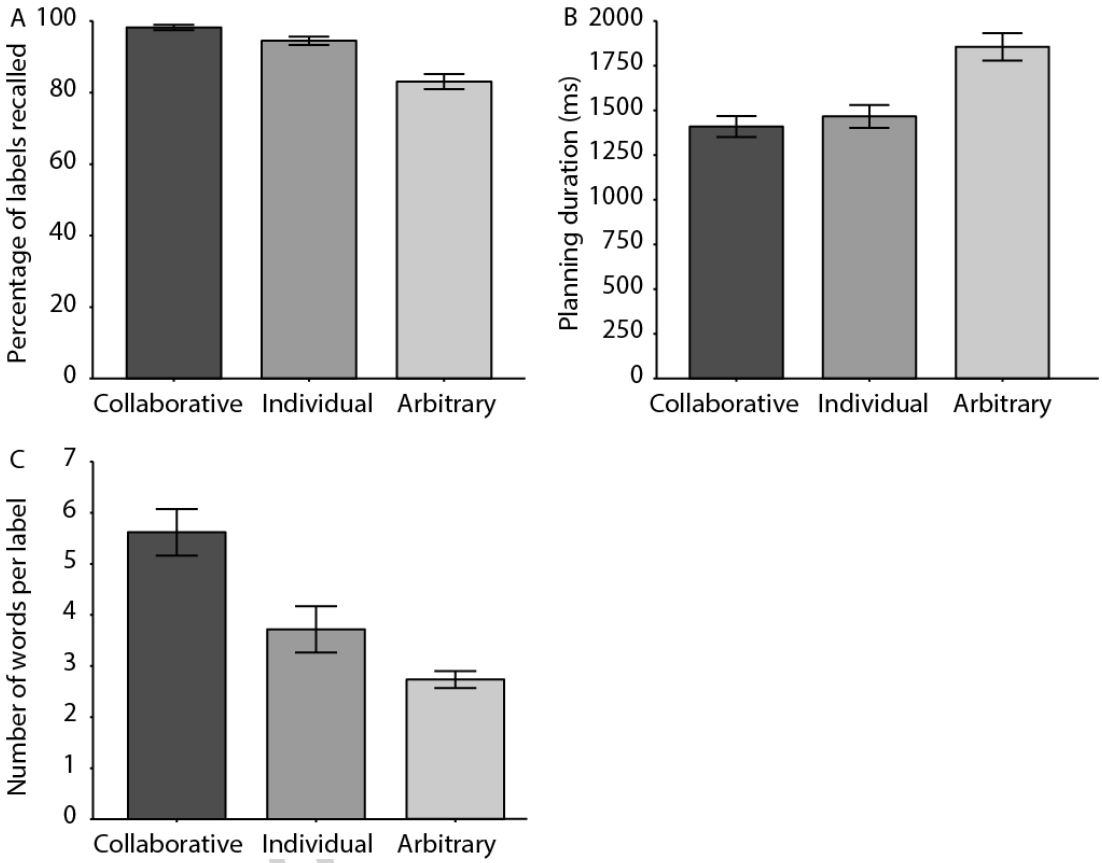


Figure 3: Behavioral results from the fMRI session: A) percentage of labels successfully recalled during the fMRI session, B) planning duration during the fMRI session (time from picture onset until the director's button press) for correct trials, and C) number of words per label averaged across the successfully recalled trials. Error bars indicate standard error of the mean.

ACCEPTED MANUSCRIPT

fMRI results

We compared the retrieval of labels learned in the collaborative condition to labels learned in the other conditions to examine the neural mechanisms of collaborative learning. In addition, we compared the retrieval of labels learned in the individual and arbitrary conditions to examine the effects of learning self-generated labels. The results of the fMRI analyses are listed in Table 1 and presented in Figures 4 and 5. The fMRI results are all from the retrieval phase of the trials modelled from picture onset.

Accepted manuscript

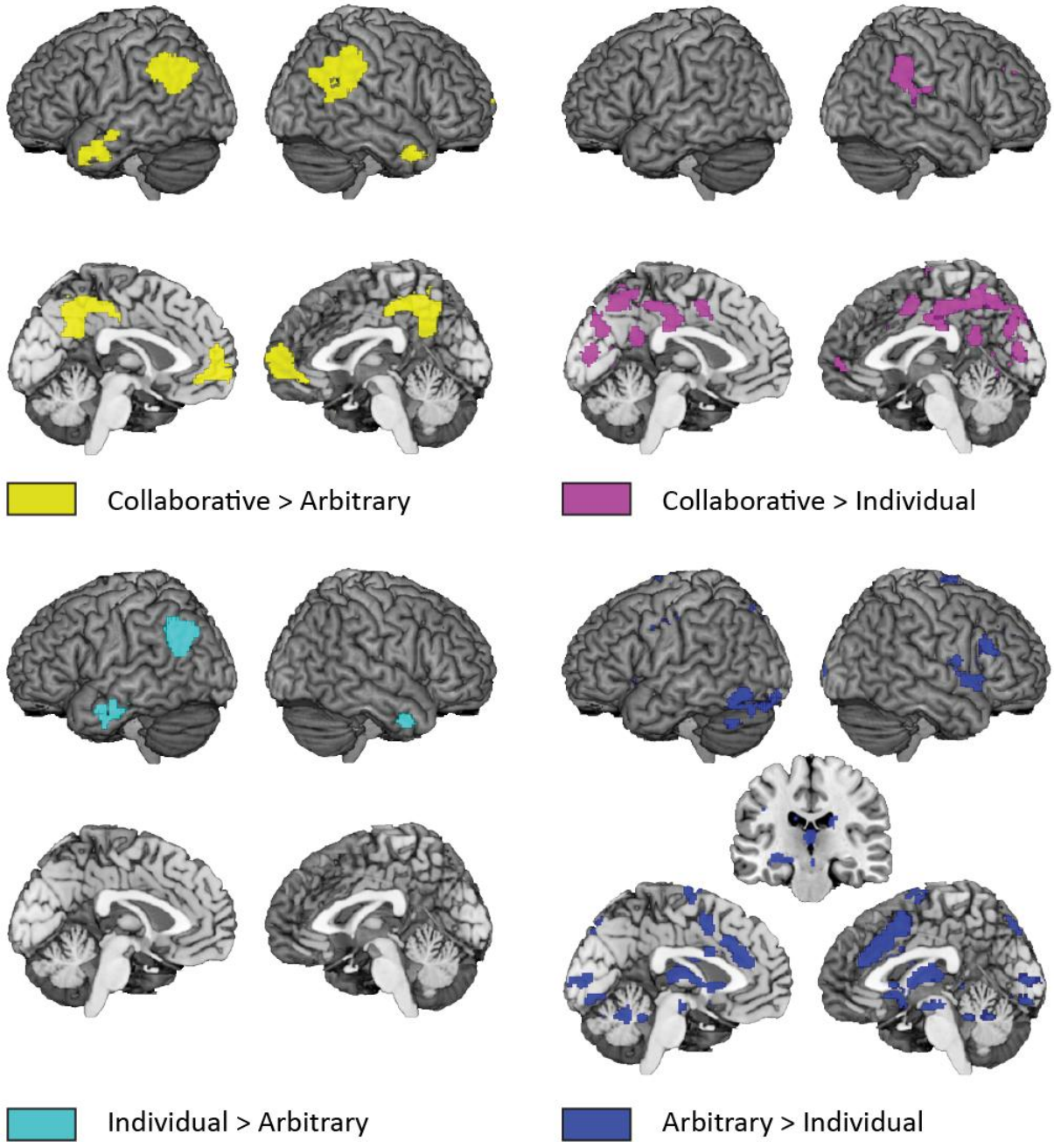


Figure 4: Brain regions showing increased activity during the successful retrieval of labels learned in different tasks by the director: the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow), the retrieval of collaboratively versus individually learned labels (violet), the retrieval of individually learned labels versus labels learned in the arbitrary task (light blue), and the

ACCEPTED MANUSCRIPT

retrieval of labels learned in the arbitrary task versus individually learned labels (dark blue). All clusters shown are significant on the cluster level and corrected for multiple comparison at p < 0.05.

Collaborative > Arbitrary. The comparison between the correctly retrieved collaborative and arbitrary conditions resulted in clusters in the precuneus, the bilateral angular gyri, medial prefrontal cortex and bilateral temporal poles (Figure 4 and 5, yellow).

Individual > Arbitrary. The contrast between the correctly retrieved individual and the arbitrary conditions revealed clusters in the left angular gyrus, and in the temporal poles extending into the middle and inferior temporal gyri (Figure 4 and 5, light blue).

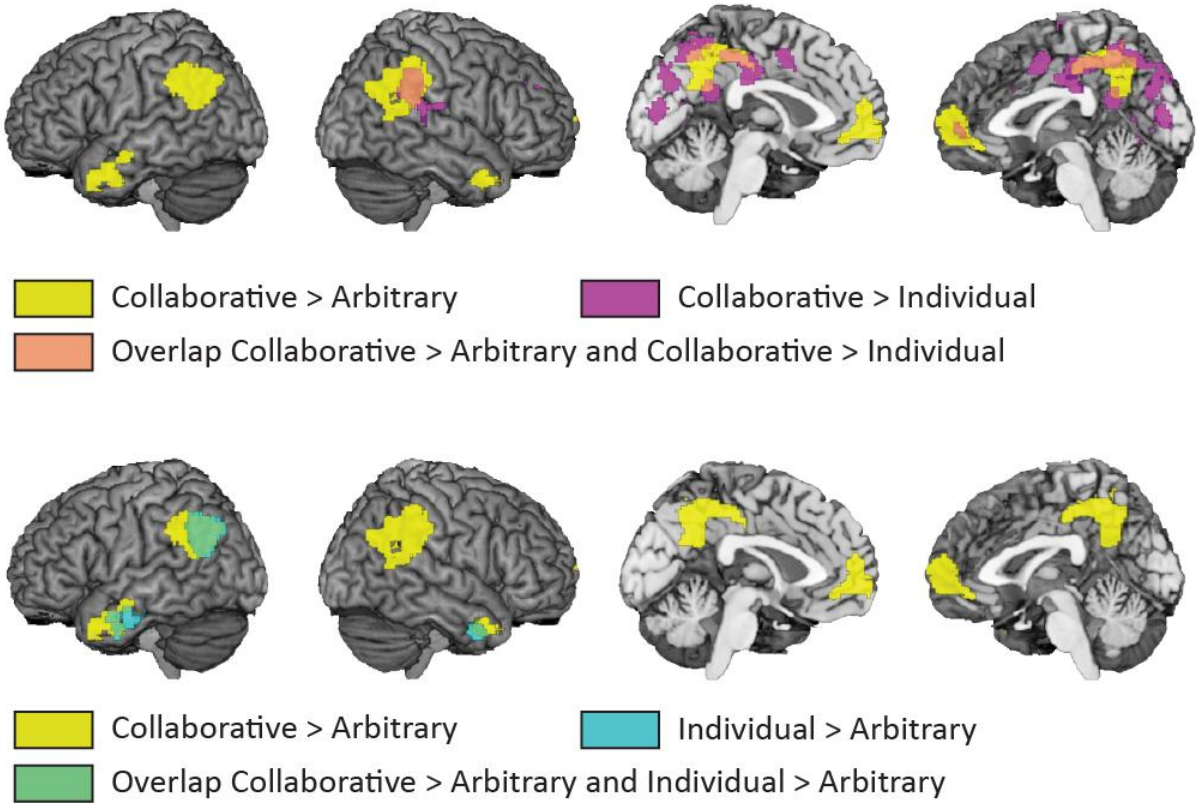


Figure 5: Overlap between the task contrasts depicted in Figure 4. Orange indicates the overlap between the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow) and the

ACCEPTED MANUSCRIPT

retrieval of collaboratively versus individually learned labels (violet). Green indicates the overlap between the retrieval of collaboratively learned labels versus labels learned in the arbitrary task (yellow) the retrieval of individually learned labels versus labels learned in the arbitrary task (light blue).

Collaborative > Individual. The contrast of main interest was between the correctly retrieved collaborative and individual conditions. This comparison revealed a series of clusters including a large cluster connecting the midcingulate cortex and precuneus, a right temporoparietal cluster, a cluster in the left putamen, bilateral middle frontal clusters and several medial frontal clusters (Figure 4 and 5, violet).

Arbitrary > Individual. The comparison between the correctly retrieved arbitrary and individual conditions resulted in a large number of clusters including the anterior cingulate cortex, bilateral caudate nucleus, left hippocampus, bilateral insula, medial frontal cortex and the bilateral calcarine gyri (Figure 4, dark blue).

Region of interest analysis. To further clarify the difference in involvement of the hippocampus in the collaborative learning context compared to the other learning contexts, we extracted parameter estimates using the MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002) from the anatomically defined left and right hippocampus defined in the AAL template (Tzourio-Mazoyer, et al., 2002). A factorial repeated-measures ANOVA on the parameter estimates revealed no significant main effects (condition: $F(2, 68) = 3.00$, $p > 0.05$; hemisphere: $F(1,34) = 3.34$, $p > 0.05$), but a significant hemisphere x condition interaction, $F(2,68) = 3.32$, $p < 0.05$. Post-hoc analyses revealed that the hippocampus was significantly more activated in the arbitrary condition relative to the individual condition, $t(34) = 3.38$, $p < 0.05$. The parameter estimates extracted from the left hippocampus are plotted in Figure 6.

Accepted manuscript

ACCEPTED MANUSCRIPT

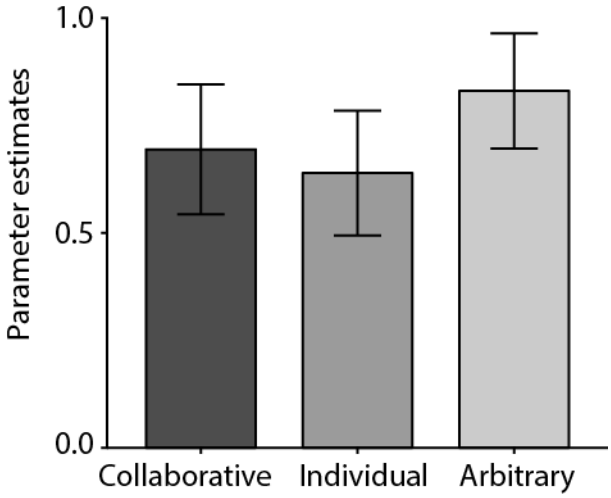


Figure 6: Parameter estimates for the collaborative, individual and arbitrary conditions extracted from the anatomical left hippocampus ROI.

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 1: Whole-brain results for the comparisons between the collaborative, individual and arbitrary conditions. Up to three local maxima are listed when a cluster has multiple peaks more than 8 mm apart.

Brain region	Cluster extent (voxels)	T value	MNI coordinates x	y	z
Collaborative > Arbitrary					
precuneus	2554	6.41	0	-46	48
right precuneus		6.12	8	-52	30
left precuneus		5.66	-12	-50	38
left inferior temporal gyrus	375	6.28	-52	-4	-32
left temporal pole		5.47	-48	10	-30
left middle temporal gyrus		5	-56	8	-26

Accepted manuscript

ACCEPTED MANUSCRIPT

right inferior parietal lobule	1546	6.2	60	-40	46
right supramarginal gyrus		5.29	64	-36	36
right angular gyrus		5.06	50	-66	28
left middle temporal gyrus	1204	5.53	-58	-56	22
left inferior parietal lobule		5.37	-60	-46	40
left supramarginal gyrus		5.14	-62	-46	32
right temporal pole	231	5.05	40	16	-34
right middle temporal gyrus		4.41	52	4	-30
right temporal pole		4.4	44	6	-38
right medial frontal gyrus	1135	4.86	6	56	4
left medial frontal gyrus		4.79	-4	54	8
right medial orbitofrontal cortex		4.62	8	50	-4
Individual > Arbitrary					
left angular gyrus	736	5.6	-54	-58	26
left inferior parietal lobule		5.51	-52	-60	40
left inferior parietal lobule		4.57	-48	-74	36
right inferior temporal gyrus	122	5.13	56	0	-34
right inferior temporal gyrus		4.14	48	6	-34
left middle temporal gyrus	196	4.34	-54	-20	-20
left inferior temporal gyrus		4.22	-50	-4	-34
left middle temporal gyrus		4.08	-60	2	-24
Collaborative > Individual					
left putamen	279	6.07	-22	10	0

Accepted manuscript

ACCEPTED MANUSCRIPT

right midcingulate cortex	5056	5.73	14	-30	40
left superior parietal lobule		5.19	-16	-60	58
right precuneus		4.88	6	-52	56
right supramarginal gyrus	514	5.61	60	-42	36
right supramarginal gyrus		4.8	66	-42	30
right superior temporal gyrus		3.83	66	-38	20
left middle frontal gyrus	121	5.58	-28	36	30
right middle frontal gyrus	162	5.11	32	38	28
anterior cingulate cortex		3.49	20	30	28
anterior cingulate cortex	42	4.71	14	20	26
right medial frontal gyrus / supplementary motor area	173	4.56	2	0	48
left superior frontal gyrus	72	4.24	-16	-10	70
left inferior frontal gyrus	42	4.24	-38	8	10
right medial frontal gyrus	49	3.89	8	56	4
right medial frontal gyrus / supplementary motor area	63	3.88	10	-14	72
right precentral gyrus		3.58	16	-24	76
right lingual gyrus	42	3.74	18	-56	-10
Arbitrary > Individual					
left inferior occipital gyrus	402	6.78	-54	-66	-16
left cerebellum		4.13	-46	-74	-22
		3.92	-30	-90	-20
left caudate nucleus	1199	6.39	-12	14	-6
left thalamus		5.26	-2	-16	12

Accepted manuscript

ACCEPTED MANUSCRIPT

right caudate nucleus		4.83	14	16	-4
left anterior cingulate cortex	1508	5.84	-6	10	28
left anterior cingulate cortex		5.38	-2	34	22
right anterior cingulate cortex		5.18	4	28	28
left cerebellum	289	5.57	-6	-56	-20
left cerebellum		4.91	-8	-72	-24
white matter	691	5.2	-12	-14	32
left insula		5.18	-32	18	-4
left insula		4.89	-34	8	20
white matter	98	5.17	-10	-32	28
white matter	241	5.11	-18	-12	-10
left hippocampus		4.86	-30	-22	-8
left amygdala		3.82	-22	-2	-10
left medial frontal gyrus / supplementary motor area	346	4.92	-8	-4	74
right medial frontal gyrus / supplementary motor area		4.89	10	-4	72
left medial frontal gyrus / supplementary motor area		3.74	0	-6	68
right middle frontal gyrus	108	4.9	32	38	28
left cerebellum	80	4.83	-30	-60	-34
right insula	950	4.8	32	24	4
right insula		4.54	52	14	-6
right insula		4.47	30	24	-4
left superior parietal lobule	1315	4.54	-30	-64	44
right calcarine gyrus		4.38	16	-98	4
left calcarine gyrus		4.37	2	-88	-10

Accepted manuscript

ACCEPTED MANUSCRIPT

white matter	52	4.5	-28	-36	28
right caudate nucleus	82	4.48	20	-20	24
right thalamus		4.18	18	-18	14
left precentral gyrus	121	4.46	-36	0	36
		3.86	-26	2	40
left precentral gyrus		3.62	-46	2	34
left cerebellum	55	4.36	-2	-46	-22
left postcentral gyrus	116	4.33	-48	-14	44
left postcentral gyrus		4.13	-42	-18	36
right rolandic operculum	42	4.22	66	0	12
left cerebellum	73	4.16	-48	-58	-34
brainstem	83	4.12	0	-14	-12
		3.75	2	-24	-12
		3.57	-8	-14	-4
right precuneus	81	4	18	-66	24
right superior occipital gyrus		3.97	26	-64	28
right calcarine gyrus		3.53	12	-78	12
	48	3.87	0	-36	4

Individual > Collaborative

right middle occipital gyrus	723	5.07	42	-90	4
right inferior occipital gyrus		4.71	40	-68	-10
right inferior occipital gyrus		4.27	30	-88	-8
right middle orbital gyrus	61	4.68	36	48	-10

Accepted manuscript

ACCEPTED MANUSCRIPT

left middle occipital gyrus	214	4.5	-32	-88	4
left inferior occipital gyrus		3.9	-44	-80	-4
left inferior occipital gyrus		3.67	-42	-72	-8
right fusiform gyrus	90	4.39	40	-18	-20
right orbitofrontal cortex	72	4.21	24	40	-16
right inferior frontal gyrus	90	4.11	46	8	26
right inferior frontal gyrus		3.98	54	10	30
Arbitrary > Collaborative					
right insula	851	8.68	34	24	-2
right inferior frontal gyrus		4.45	36	38	-12
right orbitofrontal cortex		4	34	46	-10
left inferior occipital gyrus	2496	6.98	-48	-66	-10
left inferior occipital gyrus		6.71	-48	-64	-20
left inferior occipital gyrus		6.62	-42	-62	-6
left inferior parietal lobule	733	6.9	-28	-70	46
left middle occipital gyrus		5.82	-28	-70	36
right inferior frontal gyrus	1145	6.75	50	22	28
right inferior frontal gyrus		6.24	46	28	18
right inferior frontal gyrus		5.66	50	12	32
left inferior frontal gyrus	1139	6.3	-46	22	26
left inferior frontal gyrus		6.16	-44	28	20
left precentral gyrus		5.14	-42	8	30
left cerebellum	145	5.93	-8	-72	-24
brainstem	654	5.92	-2	-14	-14

Accepted manuscript

ACCEPTED MANUSCRIPT

		4.65	10	-24	-16
		4.44	2	-20	-30
right middle occipital gyrus	1395	5.9	36	-90	10
right inferior occipital gyrus		5.3	30	-84	0
right inferior occipital gyrus		5.24	32	-90	-8
left thalamus	1743	5.9	-2	-20	12
thalamus		5.69	-4	-4	2
thalamus		5.64	10	-24	22
right superior occipital gyrus	612	5.78	30	-68	44
right superior occipital gyrus		4.05	30	-56	36
right superior medial frontal gyrus	783	5.63	4	28	46
left anterior cingulate cortex		5.01	-6	32	22
right anterior cingulate cortex		4.77	6	30	28
left insula	365	5.59	-32	22	-4
left insula		3.66	-44	14	-6
left inferior frontal gyrus		3.62	-36	36	-6
right inferior temporal gyrus	117	4.63	54	-60	-20
	106	4.49	-16	-28	26
thalamus		4.29	-8	-16	26
left fusiform gyrus	87	4.19	-32	-36	-24
left inferior frontal gyrus	43	3.87	-48	44	-2

Accepted manuscript

ACCEPTED MANUSCRIPT

Discussion

In this study, participants generated labels for abstract figures together with another person in a collaborative communication task (collaborative condition) and by themselves (individual condition), and were given pre-determined, unrelated labels to learn by themselves (arbitrary condition). They then retrieved these labels during a communication task in the MRI scanner. The analysis of the fMRI data revealed two main findings. First and most importantly, we show that collaboratively learned labels are neurally distinguishable from individually learned labels mainly in brain areas related to memory retrieval and social cognition, such as the medial prefrontal cortex, the right temporoparietal junction and the precuneus. Second, we show that the retrieval of self-generated labels as compared to unrelated, arbitrary labels engages semantic processing areas, such as the left angular gyrus and the bilateral temporal poles, while retrieving arbitrary word-picture associations is more hippocampus-dependent.

Collaboratively encoded labels are neurally distinguishable from individually encoded labels

The most direct test of the effect of collaborative encoding on memory retrieval is the comparison between the retrieval of collaboratively and individually learned labels. Labels learned in both of these conditions were self-generated, so the crucial difference between them was whether the labels were generated in a collaborative or an individual context. The results of this comparison include the right angular gyrus (often referred to as the right temporoparietal junction in the social cognition literature), the posterior cingulate cortex, the medial prefrontal cortex and the precuneus. This set of brain regions has been associated with a number of cognitive processes including theory of mind, autobiographical memory, navigation and prospection (Buckner & Carroll, 2007; Spreng, Marr & Kim, 2009; Rabin, Gilboa, Stuss, Mar & Rusenbaum, 2010). These processes all involve self-projection: mentally projecting yourself

Accepted manuscript

ACCEPTED MANUSCRIPT

into an alternative situation or perspective beyond the immediate environment (Buckner & Carroll, 2007). Of these processes involving self-projection, theory of mind and autobiographical memory seem to be the most likely candidates to explain the neural differences we found in the current study.

The main difference between the collaborative and the individual encoding tasks is the presence or absence of a matcher. Directors likely put themselves in the matchers' shoes during encoding, a process known as *theory of mind* or *mentalizing*, and may have also engaged in mentalizing to facilitate the retrieval of these jointly established labels. In line with this view, the mentalizing network has previously been found to be recruited when people perform communicative actions (e.g., Sassa et al., 2007; Willems et al., 2010) and during the encoding and retrieval of social information (e.g., Mitchell, Heatherton, & Macrae, 2002; Mitchell, Macrae, & Banaji, 2004). For example, Mitchell et al. (2004) found that subsequent memory performance correlates with activity during encoding in the medial prefrontal cortex during a social orienting task. It is interesting to note that patients with bilateral ventromedial prefrontal damage can successfully perform the tangram task (Gupta, Tranel, & Duff, 2011). A possible explanation based on our results is that this area forms part of a larger network that supports collaborative retrieval, potentially allowing for other parts of the network to compensate after damage to the medial prefrontal cortex.

Another important difference between the collaborative and the individual encoding tasks was that directors had to agree on the labels with the matchers in the collaborative task. Over the course of these interactions, directors likely formed richer episodic memory representations than in the individual task. For example, if the matcher failed to select the correct picture based on the director's initial description, the director had to come up with a novel description of the picture or provide additional detail, thus creating additional cue-target associations and facilitating retrieval (similar to the elaboration account of the testing effect proposed by Carpenter, 2009). Such prompts for elaboration were not present in the

Accepted manuscript

ACCEPTED MANUSCRIPT

individual task. Directors may have also retrieved specific social information or cues related to the matcher (e.g., voice, knowledge of a specific subject) when recalling a collaboratively learned label.

The neural differences between the retrieval of collaboratively and individually generated labels thus likely stem from the involvement of theory of mind processing and/or the encoding of more elaborate autobiographical memories for collaboratively generated labels, yet based on the current design it is difficult to pinpoint *exactly* what cognitive processes these differences reflect. Future research focusing specifically on the encoding phase or using more tightly controlled encoding paradigms should be able to determine the contributions of theory of mind and autobiographical memory to collaborative encoding and retrieval. Another important question for future studies is whether these effects are associated with collaborative retrievel more generally (e.g., a "social" retrieval strategy), or reflect partner-specific information. Healthy participants use longer descriptions when they are matched with a new partner in the tangram task, taking into account that the new matcher does not know the previously established conceptual pacts (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992). Future studies should test whether the same brain regions are activated and whether they are activated to the same extent when collaboratively generated labels are retrieved for a different addressee or in a non-communicative context.

While we also observed behavioral differences between the labels learned in the collaborative and individual contexts at retrieval, it seems unlikely that these behavioral differences can explain the observed neural differences. By including parametric modulators and modeling incorrect trials separately in the GLM of the fMRI data, we tried to account for the differences in the number of words, planning durations and accuracy that may contribute to the difference between conditions. This approach appears to have been successful, given that we did not observe increased activation for the collaborative condition relative to the individual condition in the classical, left-hemispheric temporo-

Accepted manuscript

ACCEPTED MANUSCRIPT

parietal-frontal language network where we would have otherwise expected widespread activation as a result of the longer descriptions for the collaborative condition (cf. Willems et al., 2010).

The role of semantic memory and the hippocampus

To test the effect of studying self-generated labels, we directly compared the retrieval of labels learned in the individual and arbitrary tasks. The Individual > Arbitrary comparison revealed clusters in the left angular gyrus and the bilateral temporal poles. We also found considerable overlap between the Individual > Arbitrary and Collaborative > Arbitrary contrasts in these areas (Figure 5, bottom row). In the collaborative and individual tasks, participants were free to come up with labels for the abstract figures. They probably tried to associate the abstract figures with their existing conceptual knowledge of the world. The temporal and left angular clusters we find here are therefore likely the result of retrieving semantic associations between the pictures and labels (Binder et al., 2009; Patterson, Nestor, & Rogers, 2007; Price, 2012). It is interesting to note that while we find the involvement of the left angular gyrus in both the Individual > Arbitrary and Collaborative > Arbitrary contrasts, we only found a cluster in the right angular gyrus for the Collaborative > Arbitrary comparison. These findings align with the proposed hemispheric specialization of the angular gyrus: activation in the right angular gyrus is consistently found in attention and social cognition tasks (e.g., Corbetta et al., 2000; Saxe and Wexler, 2005) , while the left angular gyrus is thought to be an important semantic hub in the brain (Binder et al., 2009).

We only found a cluster in the hippocampus in the comparison between the arbitrary and the individual conditions on the whole-brain level. However, the results of our ROI analysis (Figure 6) suggest that the reason we only found a significant cluster in this contrast is that the left hippocampus is involved to different degrees in the retrieval of labels generated in all conditions, with the arbitrary condition being most hippocampus-dependent and the individual condition being least hippocampus-dependent. While

Accepted manuscript

ACCEPTED MANUSCRIPT

patients with hippocampal amnesia can successfully perform this type of collaborative task (Duff et al., 2006), follow-up experiments by Duff and colleagues suggest that the hippocampus is involved in at least some aspects of collaborative memory. For example, patients with hippocampal amnesia are impaired in their use of definite reference in this task (Duff, Gupta, Hengst, Tranel, & Cohen, 2011) and have difficulties establishing linguistic labels for highly similar pictures (Duff et al., 2012). Interestingly, a recent study showed that patients with hippocampal amnesia are able to successfully tailor their descriptions to a new matcher (Yoon, Duff, & Brown-Schmidt, 2017). This finding suggests that establishing a link between a label and a specific conversational partner may not be hippocampus-dependent. It remains to be tested whether patients with hippocampal amnesia would perform similarly to healthy controls in the individual task we used, for which we observed the least hippocampal involvement during retrieval.

Collaboratively generated labels are better remembered

The behavioral results show that collaboratively generated labels were remembered better than labels learned in the individual and arbitrary conditions. This is an interesting finding given that the collaboratively generated labels were also on average the longest. Based on the literature on collaborative encoding and retrieval (Barber, Rajaram, & Aron, 2010; Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997), one may have expected poorer recall performance in the collaborative compared to the individual condition. One possible explanation for this finding is that the memory of interacting with another person during the learning phase may have acted as an extra cue for retrieval of the labels. In addition, the collaborative benefit we find here may be due to the fact that the collaborative task induced more elaborative processing, because directors and matchers had to agree on a set of labels for the pictures over a series of interactions. Finally, unlike previous studies on collaborative encoding, our collaborative task required participants to jointly come up with and agree on a set of labels to allow for efficient communication. This may have led participants to come up with

Accepted manuscript

ACCEPTED MANUSCRIPT

coherent labels, which promote efficient collaborative encoding (Barber, Rajaram & Paneerselvam, 2012).

Memory performance during the fMRI session was poorest for the labels that directors had learned in the arbitrary task. Even though the labels were shorter, participants' planning durations were longer and recall performace was poorer than in the other conditions. When retrieving an arbitrary label, participants could not rely on the semantics or social context in which the label was learned, making retrieval more difficult.

In the collaborative task during the behavioral practice session, we replicated the well-established observation that directors' descriptions become shorter with each round of the task. It is interesting to note that we found a similar pattern in the individual task, because most previous research using the tangram task did not include a self-generated individual condition. Hupet & Chantraine (1992) found that the number of words does not decrease over rounds when directors are instructed to describe tangram pictures that will later be presented to a matcher, suggesting that the contributions of the matcher play a crucial role in reducing the cost of the referring process. In our individual condition, directors could safely shorten their descriptions over rounds and generate labels tailored to their own idiosyncratic knowledge (cfr. Tullis & Benjamin, 2015), as they did not need to coordinate their descriptions with the matcher to establish effective labels. This resulted in considerably shorter labels in the individual condition compared to the collaborative condition and high accuracy rates during the fMRI session, although accuracy was higher in the collaborative condition.

In conclusion, the present study compared the retrieval of labels learned in collaborative, individual and arbitrary contexts. Our results show that the retrieval of collaboratively generated labels as compared to individually learned labels engages brain regions involved in social cognition and autobiographical

Accepted manuscript

ACCEPTED MANUSCRIPT

memory. This study is the first to show that collaboration during encoding can affect the neural networks involved in retrieval.

Accepted manuscript

ACCEPTED MANUSCRIPT

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268-277.

Barber, S. J., Rajaram, S., & Aron, A. (2010). When two is too many: Collaborative encoding impairs memory. *Memory & Cognition*, *38*(3), 255-264.

Barber, S. J., Rajaram, S., & Paneerselvam, B. (2012). The collaborative encoding deficit is attenuated with specific warnings. *Journal of Cognitive Psychology*, *24*(8), 929-941.

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1176.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, 19(12), 2767–2796.

Blumen, H. M., & Rajaram, S. (2008). Influence of re-exposure and retrieval disruption during group collaboration on later individual recall. *Memory*, *16*(3), 231-244.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*(9/10), 341-345.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage*, *16*(2), S497.Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, *11*(2), 49-57.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563.

Accepted manuscript

ACCEPTED MANUSCRIPT

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science*, *20*(5), 627-633.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature neuroscience*, *3*(3), 292.

Cuc, A., Koppel, J., & Hirst, W. (2007). Silence is not golden: A case for socially shared retrieval-induced forgetting. *Psychological Science*, *18*(8), 727-733.

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, 9(1), 140–6.

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The Use of Definite References Signals Declarative Memory Evidence From Patients With Hippocampal Amnesia. *Psychological Science*, *22*(5), 666-673.

Duff, M. C., Warren, D. E., Gupta, R., Vidal, J. P. B., Tranel, D. and Cohen, N. J. (2012). Teasing apart tangrams: Testing hippocampal pattern separation with a collaborative referencing paradigm. *Hippocampus, 22*, 1087–1091.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531-534.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1537), 165-176.

Gupta, R., Tranel, D., & Duff, M. C. (2012). Ventromedial prefrontal cortex damage does not impair the development and use of common ground in social interaction: implications for cognitive theory of mind. *Neuropsychologia*, *50*(1), 145-152.

Accepted manuscript

ACCEPTED MANUSCRIPT

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201-213.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, *99*(23), 15238-15243.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-Specific Effects of Social Cognition on the Neural Correlates of Subsequent Memory. *Journal of Neuroscience*, *24*(21), 4912–4917.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655-663.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976-987.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227-1235.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.

Prince, S. E., Daselaar, S. M. & Cabeza, R. (2005). Neural Correlates of Relational Memory: Successful Encoding and Retrieval of Semantic and Perceptual Associations*. Journal of Neuroscience, 25*(5), 1203–1210.

Accepted manuscript

ACCEPTED MANUSCRIPT

Rabin, J. S., Gilboa, A., Stuss, D. T., Mar, R. A., & Rosenbaum, R. S. (2010). Common and unique neural correlates of autobiographical memory and theory of mind. *Journal of Cognitive Neuroscience*, *22*(6), 1095-1111.

Rajaram, S., & Pereira-Pasarin, L. P. (2007). Collaboration can improve individual recognition memory: Evidence from immediate and delayed tests. *Psychonomic Bulletin & Review*, *14*(1), 95-100.

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on psychological science*, *5*(6), 649-663.

Read, R.C. (1965). Tangrams: 330 Puzzles. New York: Dover Publications.

Roediger, H. L., Meade, M. L., & Bergman, E. T. (2001). Social contagion of memory. *Psychonomic Bulletin & Review*, *8*(2), 365-371.

Sassa, Y., Sugiura, M., Jeong, H., Horie, K., Sato, S., & Kawashima, R. (2007). Cortical mechanism of communicative speech production. *Neuroimage*, *37*(3), 985-992.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391-1399.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9-34.

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*(1), 75-82.

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of cognitive neuroscience*, *21*(3), 489-510.

Accepted manuscript

ACCEPTED MANUSCRIPT

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and theory of systems: a meta-analysis. *Neuroimage*, *48*(3), 564-584.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: collaborative and individual processes in remembering. *Journal of experimental psychology: Learning, memory, and cognition*, *23*(5), 1160-1175.

Willems, R. M., De Boer, M., De Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8-14.

Yoon, S. O., Duff, M. C., & Brown-Schmidt, S. (2017). Learning and using knowledge about what other people do and don't know despite amnesia. *Cortex*, *94*, 164-175.

Accepted manuscript

ACCEPTED MANUSCRIPT

Highlights

- How do people learn linguistic labels in a collaborative, communicative context?
- Collaboratively and individually encoded labels are neurally distinguishable
- Neural differences in areas involved in theory of mind and autobiographical memory

Accepted manuscript