Title Title Title Title

ISBN 3-936816-XX-X

Druck: Sächsisches Druck- und Verlagshaus Direct World, Dresden

Title Titel Titel Titel Title Titel Titel Titel Title Titel Titel Titel Title Titel Titel Titel

Der Fakultät fÃ¼r
Biowissenschaften, Pharmazie und Psychologie
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium
Dr. rer. nat.

vorgelegt
von Dominic Portain
geboren am 17. Juli 1985 in Ulm

Dekan:
Prof. Dr. XXX XXX
Gutachter:
Prof. Dr. XXX XXX
Prof. Dr. PhD. XXX XXX
Prof. Dr. XXX XXX

Tag der Verteidigung: Leipzig, den 19th February 2015

Contents

Part I

Theoretical Part

Chapter 1

Introduction

1.1 Functional anatomy of language processing

1.1.1 White fiber tracts

Brain tries to save space and energy Connections cost both space and energy, so brain is largely locally organized Long-distance connections cost much more space and energy than local connections, while not transferring more information The fact that long-distance connections exist means the transferred information is especially important and/or highly compressed

1.1.2 Language network

Dorsal & Ventral streams The ventral stream uses two major long-distance fiber tracts: ECFS and UF Ventral pathway I (ECFS): STG to BA45 Ventral pathway II (UF): aSTG to FOP The dorsal stream uses two major long-distance fiber tracts: The arcuate fascicle and the superior longitudinal fascicle. Dorsal pathway I (SLF): pSTG to premotor cortex in BA6 Dorsal pathway II (AF): pSTG to BA44

1.1.3 Cortical activity

Region of interest: pSTS pSTG is involved in complex syntax "Neuroimaging studies have tried to characterise the neu- ral substrate for representing human action. Many of these studies have followed a different tradition in psychophysics and developmental psychology of investigating the percep- tion of "biological motion" - that is, the characteristic articu- lated motion of chordate animal bodies (e.g. Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001; Grossman & Blake, 2002; Beauchamp, Lee, Haxby, & Martin, 2002; Pelphrey, Mitchell, McKeown, Goldstein, Allison, & McCarthy, 2003) or body and face parts (e.g. Hoffman & Haxby, 2000; Hooker et al. 2003; Kilts et al. 2003; Pelphrey et al., 2003). Biologi- cal motion can also be perceived from the relative motion of just a few dots (“point-light walkers“, Johansson, 1973); if the dots are spatially or temporally rearranged, the percept is destroyed. These neuroimaging studies suggest that one brain region, the posterior superior temporal sulcus, is par- ticularly involved in the representation of biological motion. Two sets of recent neuroimaging data suggest that the role of the posterior superior temporal sulcus (pSTS) may extend beyond a response to biological motion, to more ab- stract representations of intentional action. First, Castelli, Happe, Frith, & Frith, (2000) and Schultz, Grelotti, Klin, Kleinman, Van der Gaag, Marois, & Skudlarski, (2003) reported that a region of the pSTS showed a significantly higher response to animations of moving geometric shapes that depicted complex social interactions than to anima- tions depicting inanimate motion. Second, using movies of human actors engaged in structured goal-directed ac- tions (e.g. cleaning the kitchen), Zacks, Braver, Sheridan, Donaldson, Snyder, Ollinger, Buckner, & Raichle, (2001) found that activity in the pSTS was enhanced when the agent switched from one action to another, suggesting that this region encodes the goal-structure of actions. Both of these results are consistent with a role for a region of pSTS cor- tex in representing intentional action, and not just biological motion.“ (doi:10.1016/j.neuropsychologia.2004.04.015)

Accurate role in syntactic processing isn’t clear yet Utilizing dorsal pathway I and II for researching pSTG Paradigm of choice: Social interaction, Object-/Subject-relative clauses Previous findings: pSTG and IFG (BA44, 45, 47 + FOP) are more active in object-relative clauses

1.2 Developmental aspects

Kids don’t have the AF tract yet Previous findings: kids rely more on the ventral pathway Ventral processing involves BA45 (no condition effect in adults) Processing is more vulnerable to bias (semantic crosstalk) Behavioral data is worse too

1.3 Research questions

Replication of EEG/fMRI results with MEG? Which cortical regions are involved in the conditional effect? Is semantic content relevant? How long does each processing stage take? In which order do the steps take place? Can we see a different pathway in kids?

1.3.1 Hypotheses

Condition effect mainly in pSTG, BA44 (adults), BA45 (kids) Kids: worse performance than adults Kids: less involvement of pSTG Adults vs kids: Dorsal II vs. Ventral II

1.4 Choice of measurement methods

1.4.1 Acquisition

MEG: Very high sampling rate very little spatial distortion from the head Low signal-to-noise ratio

1.4.2 Forward models

Based on individual high-resolution anatomical data 1-layer BEM: Robust creation, good software support Quick and semi-automated process

1.4.3 Localization

Distributed source model: models a ton of unwanted activity -> high SNR for actual ROI sLORETA: no overlap effect at the boundaries, high ROI specificity

1.4.4 Information transfer

Meaning of information transfer modelling vs. model-free, linear vs. nonlinear Transfer Entropy is suited best for highly complex, nonlinear timescale data Potential pitfall: Volume conduction (can also come from bad localization) TrenTOOL can correct volume conduction

Chapter 2

Pilot Study

Experimental paradigm The pilot study was designed with two goals: First, to find any strong inherent bias in the stimuli material. Especially the choice of animal pairings (predators, prey and insects mixed with each other) was a potential confounding effect. It was also unclear if the different types of actions (predatory and antropomorphic/social) could introduce a response bias.

Second, to determine whether 10-year-old children are the correct age bracket for examining our research questions. We expected a lower performance than adults, but better or equal performance than from younger children. Especially important was to test the ability of our subjects to answer questions in object-relative clause correctly.

We employed a repeated measures factorial design with one factor: syntax order. The two conditions of this factor differed in the use of either subject-relative or object-relative clauses.

Participants 21 children ([9 female]) were recruited from the internal participants database. Subjects were selected if they spoke German as native language, if their language development was unremarkable, if their handedness score was above 70 and if their medical history was free of cognitive abnormalities. Children were aged between [10y0m] and [10y11m] and described as right-handed by their parents. Parents gave written informed consent and were compensated with 7,50€. Children agreed to participate in the study and were compensated with a toy at an approximate value of 12€. All experimental procedures were approved by the University of Leipzig Ethical Review Board.

Task and Stimuli The session consisted of two sections: a tutorial and the main experiment. The tutorial section introduced the physical interface and 36 practice trials. The main section consisted of 18 clusters with 12 trials each (216 total). All trials were randomized at the beginning of the experiment. There was an exception to the random order: no two identical images were presented after another. Subjects were shown a feedback screen at the end of each cluster.

Each trial started by showing a set of visual stimuli. 200ms later, a spoken sentence started playing. Possible responses included pressing the left, right or the skip button. The trial ended with an auditory and visual feedback.

A set of visual stimuli consisted of a two side-by-side images on black background. Each image depicted two cartoon animals on a white background. In one picture, one of the animals performed a social action on the other animal. In the other picture, the roles were reversed. Subjects needed to identify the picture that answered the question correctly.

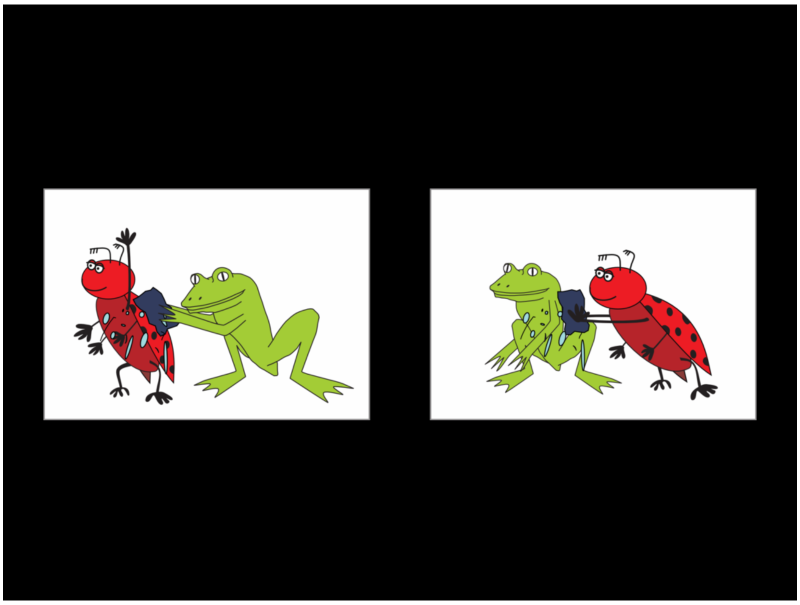

Figure 2.1: Example of a typical visual stimulus before the response

The spoken sentences were always posed as questions in order to elicit an immediate response. Each question was posed in a right-branching structure (see table 2.1 for an example). Object-relative clauses and subject-relative clauses were presented in random order at a ratio of 2.5:1. Questions were voiced in a natural, child-directed tone by a professional native speaker. The question consisted of an actor, a recipient and the verb that described their interaction. The performed action was either painting, pushing, combing, washing, pulling or catching. The actor and recipient were selected randomly from 12 different races: Lion, rabbit, wolf, bird, fox, hedgehog, dog, tiger, ape, ladybug, bear and frog. No two identical races appeared in the same picture.

Original	Wo	ist	der	Käfer,	den	der	Frosch	wäscht?
Translated	Where	is	the	bug_{OBJ},	who_{ACC}	the	frog_{SUBJ}	washes?
Word index	1	2	3	4	5	6	7	8

Table 2.1: Example stimulus sentence. Top: original spelling in German. Middle: Literal translation in English. Bottom: Word index within the sentence

Immediately after each response, an icon appeared below the two animal pairs. A green checkmark, a diagonal red cross and a yellow skip symbol signified a correct response, an incorrect response and an invalid trial, respectively. The trial feedback screen was presented for a random interval between 400ms and 800ms.

This experiment created a tradeoff between speed and accuracy. To encourage a high level of attention and a high number of usable trials, a feedback screen was displayed at the end of each cluster. On this screen, two bar graphs visualized response speed and accuracy during the preceding cluster.

Visual and auditory stimuli were produced by a computer running the software package Presentation (Neurobehavioral Systems, Inc., version [14.6]). Video signal was displayed by a 17-inch TFT display at a distance of approximately 80cm. Sound was played with a pair of semi-open headphones.

Analysis Behavioral data were analyzed with Matlab (version 2014a). Response accuracy was evaluated for the case that subjects responded randomly. For this purpose, accuracy was compared to the outcome of a random sequence of binary events. We established the alpha = 0.01 confidence interval for the percentage of correct trials ($\frac{k}{n}$) that could be answered correctly purely by chance. This calculation was implemented using the binomial fit method in Matlab: binofit(k, n, alpha). If the upper confidence interval of this calculation exceeded the subject-specific accuracy, the subject was removed from further analysis. Two subjects failed to exceed chance level performance, leaving 19 subjects for the analysis.

Two types of behavioral data were analyzed for group and condition effects: response time (RT) and response accuracy (RA). Response time was measured at the condition onset, i.e. at the “d“ sound of the sixth word. Trials were omitted when the subject skipped or answered them incorrectly, or responded earlier than the cue. Accuracy was calculated by dividing the amount of correct trials by the amount of total trials for each subject.

A Shapiro-Wilk test was used to test for normal-distributed residuals. Accuracy passed this test at a p = 0.01 significance level. The impact of the syntactic condition on response accuracy was determined with a T-test.

Any unintended bias on response time was determined by splitting RT from each subject into two groups along one of five impact factors. The two groups were then compared with a T-test. The five grouping factors were response side (left / right), condition (object-relative / subject-relative), the race of the actor (12) and recipient (12) and the peformed action.

Results Subjects responded after a median delay of 2.0s (quickest 5%: 1.0s, slowest 5%: 5.2s). The median accuracy was 70% (worst 5%: 62%, best 5%: 94%). The syntactic condition had a highly significant effect on accuracy ($p < 0.001$, $t(18) = -14.0$). No factor had a significant effect on response time ($p > 0.1$, $F < 1.5$).

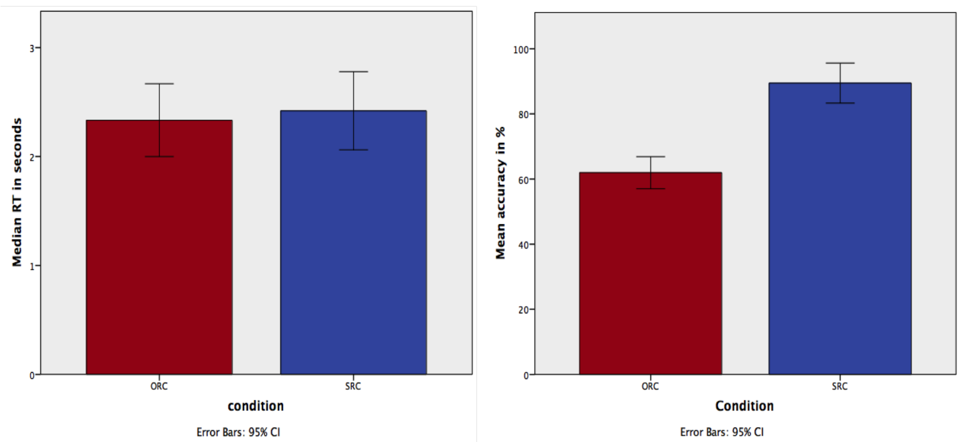

Figure 2.2: Chart of the grand average performances for subject-relative (blue) and object-relative (red) clauses. Left: Median response time from conditional onset. Right: Average response accuracy.

Post-hoc tests were conducted to reveal condition-specific accuracy values. The average accuracy for responses to subject-relative clauses was much higher than to object-relative clauses (93% and 64%, respectively). Due to the proximity to a 50% chance level, we repeated the binomial fit test of individual accuracy, exclusively with responses to object-relative clauses. 8 of 19 subjects failed this test at a $p = 0.01$ significance level. After removing these subjects from the analysis, selected tests were repeated.

Median response times remained unchanged for both conditions. Overall accuracy in the remaining subjects improved slightly (95% for subject-relative clauses and 68% for object-relative clauses). The difference in accuracy between conditions weakened slightly ($p < 0.001$, $t(10) = -9.4$).

Discussion The first goal of this pilot study was to establish an unbiased test paradigm. The variation in grammatical elements had no undesired impact on response times. Syntax conditions produced a strong effect in both performance metrics. These findings support the current stimulus setup for use during MEG measurements.

The second goal of the pilot study was to determine if 10-year old children were suitable subjects for the designed task. Accuracy levels were comparable with the findings of similar experiments. Due to different trigger points and sentence lengths, comparisons of response times couldn't be made directly and were instead limited to comparisons of effect size. We start by comparing our results to our spiritual predecessor, the study by [2.1]. They found no significant conditional impact on response time in the 9-10-year age bracket. In stark contrast to our results (93% and 64% for subject- and object-relative clauses), their subjects were not influenced by a condition effect, with accuracy levels of 94% in both conditions. The good performance was met with surprise and speculation that semantic cues may have helped with sentence comprehension. This speculation was supported by their fMRI findings, which indicated that children relied heavily on semantic-centric processing areas, rather than on pure syntax-related processing areas that adults use. Our setup didn't include semantic cues, which puts our results more in line with other infant studies.

[2.2], for instance, measured repetition performance in 3- and 4-year-old children. The study presented German subject-relative and object-relative clauses with two interacting people in third person, similar to our setup. Their subjects performed with an accuracy of just 5% (3 years) and 26% (4 years) for object-relative clauses. Subject-relative clauses were performed with 13% and 31% accuracy, respectably. Note that these ratings represent accurate verbal repetition, and don't need to be corrected for chance level performance.

[2.3] found more accurate responses to portuguese right-branching subject- and object-relative clauses. Subject-relative clauses were correctly answered 23%, 50%, 83% and 80% of the time (in 3-, 4-, 5- and 6-year old children, respectably). Object-relative clauses reached only slightly lower accuracy levels of 18%, 40%, 75% and 68%, respectably. They employed a more user-friendly approach by requiring the children to act out the posed sentences with toy animals. Because the authors deliberately removed as many processing constraints as possible, these accuracy levels can be considered upper performance limits on these age brackets.

6 to 8 year old children solved syntactic problems with higher complexity in the study by [2.4]. The experimental design varied two syntactical factors and one contextual factor. One syntactical factor, “question“, varied English object- and subject-relative clauses, coinciding with our setup. Object-first and subject-first clauses were responded with an accuracy of 64% and 83%, respectably.

Response time and accuracy performance indicates that the 10-year age bracket was successful in selecting subjects with an incomplete dorsal tract II. Compared to typical adults, our subjects performed considerably worse in both accuracy and response time. Compared to younger children, our subjects performed condsiderably better in subject-relative clauses.

42% of our subjects failed to perform better than chance level when the task required a fully-developed AF. We have no sufficient reason to assume that these subjects were using a random-button strategy. If anything, the weakened condition effect indicates that these performances were due to honest mistakes. Hence, there is no reason to exclude subjects with chance-level performances to object-relative clauses.

Finally, the pilot study revealed a few design flaws as well.

First, sentences differed systematically between conditions even before the intended condition cue. The syntactical condition reverses the order of pronouns, creating an opposition between "der den" and "den der". To prevent confounding effects from different content, the initial sentence fragment needs to be identical at least across conditions, and, preferably, across trials as well. Ideally, the distance between the end of this identical sentence fragment and the conditional cue point should be as short and invariant as possible.

Second, our sentence structure contained a theoretical loophole. With sufficient time and wit, subjects could develop an alternative strategy that doesn't require syntactic processing of the whole sentence. The alternative strategy exploits the fact that the minimum information for a correct decision is already available at the fifth word (see table **??**). Subjects only needed to complete three sequential steps: First, attending only to the left of the two pictures. Second, waiting until the fourth word is spoken. If the mentioned animal was displayed as actor, the left button would be correct and vice versa. Third, using the fifth word to execute the previous or the reverse button mapping. If the mentioned word was a "der", the previously correct button remained correct and could be pressed immediately. If the mentioned word was a "den", the previously wrong button had to be pressed for a correct result. This strategy could potentially reduce the task into a simple series of motor preparation and pattern matching. Complex syntactic processing and, presumably, the use of pSTS and dorsal pathway II would be circumvented. Employing this strategy could create suspicious behavioral results in the form of systematically reduced and less varied response times. This flaw was the main reason for the redesign of stimuli for the main study. Other minor flaws were also corrected for improved validity.

Chapter 3

Methods

3.1 Participants and stimuli

3.1.1 Participants

18 children and 22 adults were recruited from the internal participants database. Subjects were selected if they spoke German as native language, if their language development was unremarkable, if their handedness score was above 70, if they fulfilled the prerequisites for MRI scans, and if their medical history was free of cognitive abnormalities. 4 children and 4 adults dropped out inbetween sessions of the study. Two children were excluded from the analysis because their behavioral performance was at chance level. 12 children (5 female) and 18 adults (9 female) were left for the subsequent analysis. Children were aged between 9y11m and 10y9m and described as right-handed by their parents. Adults were aged between 22 and 33 years and scored between 73 to 100 (median: 95) on the [LQ] handedness test. The point of reference for these ages is the time of the MEG session. Parents gave written informed consent and were compensated with 40€for the MEG session and 7,50€for the MRI session. Children agreed to participate in the study and were compensated with a 10€gift voucher for each session. Adult participants were compensated with 20€. All experimental procedures were approved by the University of Leipzig Ethical Review Board.

3.1.2 Task

The study consisted of two sessions: an anatomical MRI acquisition (duration: 50 minutes) and an interactive magnetoencephalographic measurement (typical duration: 90 minutes). Since they took place in two different locations, there was a delay (median: 98 days, maximum: 243 days) between the two sessions.

The MRI session is described in detail in [section 3.2.2].

The MEG session consisted of two sections: a tutorial section and a main section. Generally, the participants were seated on a comfortable chair. Visual and auditory stimuli were produced by a computer running Presentation (version 14.0) at 60Hz refresh rate and 1024x768 resolution. Video signal was routed through a video splitter MSV1235 into a Panasonic PT-D7700 projector. Audio signals were generated by a Soundblaster Audigy 2 ZS [SB0350]. An audio amplifier (Compumedics, Hamburg, Germany) drove a pair of TIP-300 loudspeakers (Nicolet, Biomedical Madison, WI, U.S.A.). Sound was routed through a pair of plastic tubes (50cm length, approx. delay of 1.6ms) Sound arrived in the subjects' ears via ER3-14A/B earplugs (Etymotic Research Inc., Elk Grove Village IL, U.S.A.). Loudness was calibrated for 50db above the subjects' hearing threshold.

MEG tutorial section First, the tutorial section described the usage of the interface.

Second, subjects needed to respond to an example stimulus with the spoken sentence written out below the screen.

Third, three example trials followed without the written sentence.

Fourth, an artificially incomprehensible sentence was presented together with otherwise innocuous visual stimuli.

When subjects pressed either response button instead of skipping the trial, they were instructed with the skip function.

Finally, a series of randomized tutorial trials followed. When subjects showed behavioral proficiency of the task, the tutorial ended automatically. Two thresholds for proficiency were possible: either an average response time below 3000ms and an accuracy score above 80%, or an accuracy score above 88%. Otherwise the tutorial ended after 36 trials.

MEG main section The main section was used for MEG acquisition and consisted of 304 trials grouped in two blocks. There was a scheduled break between the blocks (usually 1-2 minutes) which included interaction with the research assistant. Subject-specific trial randomization was performed before the task. All stimuli-related randomization tasks were implemented with a time-seeded Mersenne-Twister approach in Python 2.7. Randomization contained two exceptions: neither the same image nor the same sentence could be played twice in a row. Each block consisted of 8 clusters. Subjects were shown a feedback screen at the end of each cluster, summarizing their performance throughout the recent cluster. Since manual intervention was necessary to proceed to the next cluster, subjects frequently used this opportunity for a tiny break (typically 5-20 seconds). Each cluster consisted of 19 trials.

Structure of a single trial Each trial started by showing two pictures side-by-side. In one picture, one of the animals performs a social action on the other animal. In the other picture, the roles are reversed. 10ms later, the spoken question started playing. The subject could respond by pressing one of the direction buttons or the skip button. There were two direction buttons, left or right, signifying that the left or right image contained the answer to the question. The skip button was used to mark the trial as invalid for further analysis, and excluded the trial from performance feedback. This response was the correct choice when the subject was distracted or failed to comprehend the question immediately. This opened a minor pitfall: subjects could have gotten perfect scores by just pressing the skip button each time. Fortunately, none of the subjects discovered this opportunity. The trial ended with an auditory and visual feedback.

3.1.3 Visual stimuli

Character motivation A set of visual stimuli consisted of a two side-by-side images on black background. Each image depicted two different animals on a white background. I selected selected social activities that were only plausible for antropomorphised characters. Antropomorphization includes the use of their front limbs for object manipulation and standing on their hindlegs. These measures are introduced to prevent associations with real-world animalistic behavior. For example, a lion “catching“ a monkey could resemble predatory behavior. This association with chasing and killing would introduce a semantic bias against the reverse interaction: a real-life monkey “catching“ a lion is much more implausible than the reverse. To further detract from a naturalistic view, the animals were represented in a cartoon style. To prevent unnecessary stress on this interpretation, I only selected animals whose real-life counterparts were approximately equally sized.

Image components were adapted with permission and kind advice from [5]. Modifications were performed with Inkscape.

Trial feedback Immediately after each response, an icon appeared below one of the two displayed animal pairs. The presented side was determined by the subject’s response. In the case of the skip button, the icon appeared at the same height as the others, but in the middle of the screen. A green checkmark, a diagonal red cross and a yellow skip symbol signified a correct response, an incorrect response and an invalid trial, respectively. The trial feedback screen was presented for a random interval between 400ms and 800ms.

Cluster feedback In this experiment, there was an obvious tradeoff between speed and accuracy. To encourage a high level of attention and a high number of usable trials, two bar graphs visualized performance speed and accuracy (see Fig. 3.4). In order to maximize the

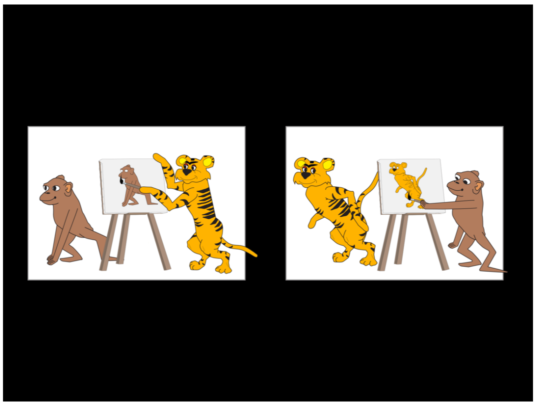

Figure 3.1: A typical visual stimulus, featuring two pairs of animals.

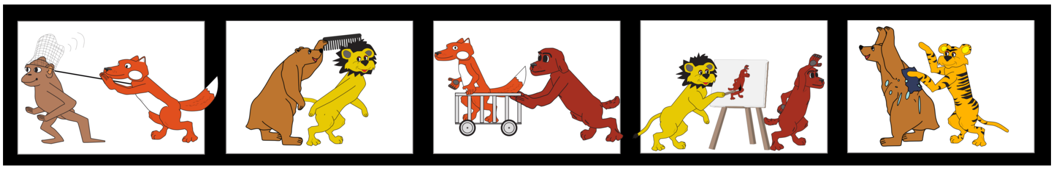

Figure 3.2: Illustrations of all five animals performing their social activities. From left to right: catching, combing, pushing, painting and washing

amount of usable trials for further analysis, the visualization valued accuracy much more than response time (see Fig. 3.5).

3.1.4 Auditory stimuli

Sentence content Each pair of images was presented with a spoken question. The question format fit well with the stimulus-response paradigm, and allowed the sentences to be identical until the conditional article (“den“ or “der“) appeared. Syntactically, the sentences used an equal number of subject-relative and object-relative clauses. In order to minimize confounding effects, these two conditions were designed to show as little auditory distinction as possible. The structure of the final sentences is displayed in table 3.1.

Tutorial sentences During the pilot study, children often assumed that the every sentence was a subject-relative construction, miscategorizing “den“ for “der“. Tutorial sentences made the two animal nouns explicit, so that all sentences were structured in the format “Where is the monkey that is caught by the dog?“. While this setup was creating strong

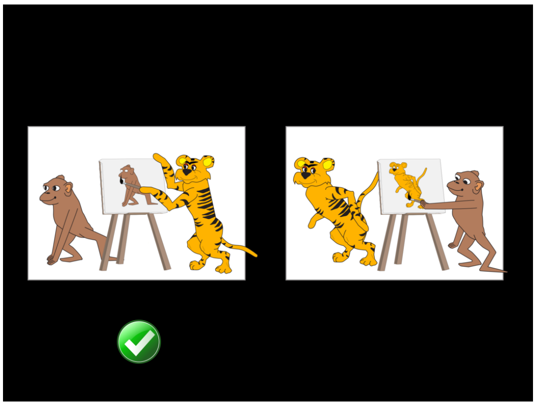

Figure 3.3: The visual feedback to a correct response.

Original	Wo	ist	das	Tier,	das	der	Tiger	malt?
Translated	Where	is	the	animal$_{OBJ}$,	which	the$_{NOM}$	tiger$_{SUBJ}$	paints?
Word index	1	2	3	4	5	6	7	8

Table 3.1: Example stimulus sentence. Top: original spelling in German. Middle: Literal translation in English. Bottom: Word index within the sentence.

auditory differences, it was easier to comprehend. If the children didn't notice the difference by the eighth tutorial trial, the research assistant repeated the question with an exaggerated "den" pronounciation.

Audio format Sentences were spoken by a professional female native speaker in an unisonous and moderately child-directed prosody. Recording and playback was performed at a sampling rate of 44100Hz with one channel. Loudness of each sentence was normalized. Overall loudness was adjusted to [60db] above each subject's individual hearing threshold.

Trial feedback Immediately after each response, one of two short sounds played. The sounds were extracted from Microsoft Windows XP. The "external device plugged in" icon (two bell sounds in ascending tone) and the "external device removed" icon (two bell sounds in descending tone) represented correct and incorrect responses, respectively. No sound was played after the skip button.

Figure 3.4: An ideal cluster feedback screen that appears after completing 19 trials. Upper bar: speed, lower bar: accuracy. Bottom right: indicator to press the skip button to advance

3.2 Data acquisition

3.2.1 MEG

MEG data were collected with an Elekta Neuromag VectorView® MEG scanner in Bennewitz, at the department of Magnetoencephalography, Institute for Cognitive and Brain sciences, Leipzig, Germany. The scanner comprised 306 MEG-channel sensors (102 magnetometers, 204 planar gradiometers). Sensors were tuned prior to each MEG recording session to limit noise levels to approximately 2.5 fT/cm. Sensors that became very noisy during a recording block would be individually re-tuned at the next inter-block break, using the Neuromag automatized heating process or by eye, as necessary. Continuous MEG data were recorded at 1000 Hz sampling rate (0.3-330 Hz bandpass filter).

Prior to data acquisition, all metal and other potential sources of electromagnetic interference were removed from participants. Quality of recording was confirmed by visual inspection of a live view of MEG recording before each session without the subject present. Electro-oculogram (EOG) and electrocardiogram (ECG) time-series were recorded simultaneously with MEG to track potential noise sources and artifacts. Five head position indicator (HPI) coils were attached to the participant's forehead and a Polhemus stylus and digitizer device were used to record the locations of fiducial points (right and left pre-auricular points (RPA, LPA) and nasion), the HPI coils, and between 150 and 200 extra digitizer points on the head surface. Prior to the recording of each stimulus block, head location in the scanner was measured with an automatic process that detected the coils. Continuous HPI recorded any head movements during data acquisition.

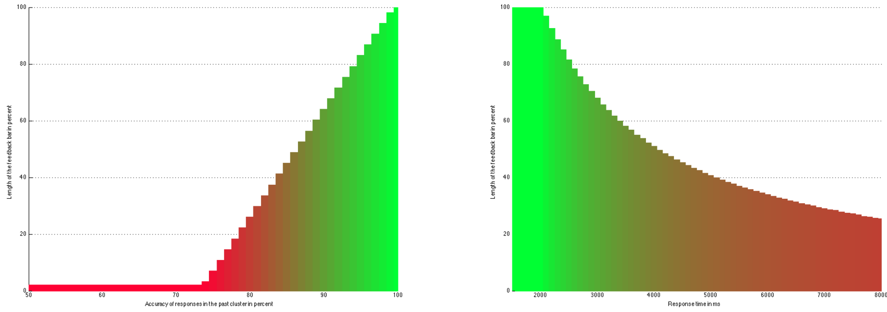

Figure 3.5: Relation between performance and displayed feedback bars. Performance (left: RA, right: RT) is drawn along the X-axis, and length of the bars in % is drawn along the Y-Axis.

3.2.2 MRI

Anatomical magnetic resonance imaging (aMRI) data were collected with a 3.0 Tesla TIM Trio scanner, located at the Max-Planck-Institute for Cognitive and Brain sciences. Two scans were acquired from each participant in one session: A T1-weighted scan and a T2-weighted scan. The T1-weighted scan used the magnetization-prepared rapid gradient echo (MPRAGE, [6]) sequence (flip angle = 9°, TR/TE/TI = $2300ms/2.96ms/900ms$). This scan was oriented transverse (176 slices) with an isotropic resolution of 1mm. The T2-weighted scan used the SPACE sequence by [7] (flip angle = 120°, TR/TE = $3200ms/402ms$). This scan was oriented transverse (176 slices at 1mm) with an inplane resolution of 0.5mm x 0.5mm. All scans used a 32-channel head coil for the acquisition.

3.3 Data analysis

Data were preprocessed with the three software packages: Elekta Neuromag® MaxFilter (version 2.2), Matlab (version 2014a) and MNE-Python (version 0.8.6).

3.3.1 Behavioral data

Two types of behavioral data were analyzed for group and condition effects: response time (RT) and response accuracy (RA). Response time was measured at the condition onset, i.e. at the "d" sound of "den" or "der" (in the subject-relative clause or the object-relative clause, respectively). Trials were omitted when the subject skipped or answered them incorrectly. Trials were also omitted if the response took longer than 4000ms. This procedure removed 11.1% of the childrens' trials, and 2.5% of the adults' trials.

RT and RA were determined for each subject separately from the remaining trials. Both metrics were tested for the requirements for an analysis of variance (ANOVA). Normality of the residuals was tested with a Shapiro-Wilk test [swtest], implemented in Matlab. Equality of variances was tested with a Levene [levtest] test, implemented in SPSS. RA data failed the normality test. To include RA data in the following analysis, they were transformed to fit a normal distribution. This transformation was accomplished with the inverted sigmoid function:

$$\hat{a} = -log(\frac{1}{a} - 1)$$

All results from the ANOVA were transformed back into milisecond space with the sigmoid function:

$$r = \frac{1}{1 + e^{-\hat{r}}}$$

3.3.2 Sensor-space activity

Preprocessing and HPI correction Signal-space separation (in MaxFilter) was used to reduce noise in the data by suppressing magnetic interference coming from outside and inside the sensory array. MEG recordings were corrected for HPI movements, and co-registered across blocks to the inital head position for each individual (processing in MaxFilter). Data were then subjected to a 0.4Hz FIR highpass filter (Hamming window design, 4367 coefficients, -130db damping at 0Hz, processing in Matlab) to remove slow trends.

Artifact removal MEG channels with abnormally high noise levels as identified by visual inspection were rejected from further analysis. A median of 1 channel (maximum: 3 channels) was removed. The resulting pre-processed data contained major artifacts from spontaneous channel jumps, electrocardiographic (ECG) activity and electrooculographic (EOG) activity. Jump amplitudes were detected by selecting peaks in the z-transformed continuous data that exceeded a threshold of 12 standard deviations. Segments of 2 seconds in the pre-processed continuous data were rejected if any magnitude channel exceeded an amplitude of $6 \cdot 10^{-12}T$ (gradiometer channels: $4 \cdot 10^{-12}\frac{T}{cm}$). Continuous data were then decomposed into independent components (ICA) that explained 99% of the variance. Components that correlated with EOG or ECG channels were removed with the MNE methods preprocessing.ica_find_ecg_events() and preprocessing.ica_find_eog_events(), respectively. ICA-based correction removed an average of 2.1 components per subject and block (minimum: 1, maximum: 4). The remaining ICA components were used to reconstruct continuous data.

Epoching The main trigger was set at the condition onset (described in [3.3.1]). Epochs were created between 1000ms before and 4000ms after the main trigger. An epoch was rejected if the trial was skipped, or answered too slow (more than 4000ms) or answered incorrectly. This procedure yielded an average of [] trials in children and [] trials in adults. Data were filtered before epoching with a 45Hz FIR lowpass for visualization purposes only.

Establishing time windows of interest The condition effect was used to determine suitable time windows. This bootstrapping strategy typically causes overfitting. This problem was resolved with a cluster-level permutation comparison. Since the group had a strong impact on RT (see [4.1.1]), effective time windows were estimated separately for children and adults.

First, extracted trials were split into a 4 parts (2 groups x 2 conditions).

Second, all trials from the selected part were pooled.

Third, sensor data was pooled by calculating the mean from one of four different sensor groups. General locations were selected according to previous literature findings. The equivalent sensor selections were determined by the MNE function read_selection, and are visualized in 3.6.

Finally, clusters were computed by the MNE function stats.permutation_cluster_test [8]. The function was run with 2500 permutations, and an t-threshold of 1.0.

For visualization purposes, grand average activity was also calculated for each sensor group and condition, separately for children and adults.

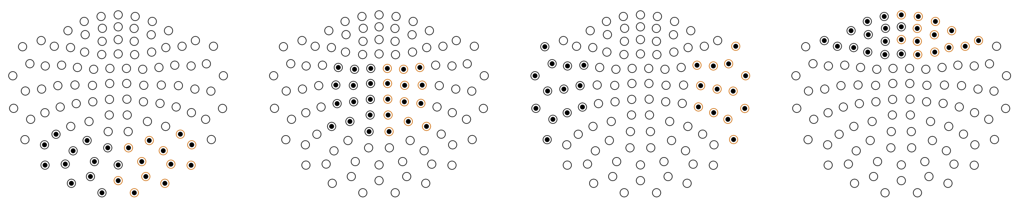

Figure 3.6: Selected channels for each sensor location. From left to right: occipital, parietal, temporal, frontal. The right hemisphere (in red) is also depicted on the right side in each illustration.

3.3.3 Source space activity

Anatomical preprocessing Cortical reconstruction and volumetric segmentation was performed with the Freesurfer image analysis suite, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The technical details of these procedures are described in prior publications (Dale et al., 1999; Dale and Sereno, 1993; Fischl and Dale, 2000; Fischl et al., 2001; Fischl et al., 2002; Fischl et al., 2004a; Fischl et al., 1999a; Fischl et al., 1999b; Fischl et al., 2004b; Han et al., 2006; Jovicich et al., 2006; Segonne et al., 2004, Reuter et al. 2010, Reuter et al. 2012). We followed the recommended processing pipeline ("recon-all"), with three optional functions.

First, the option "-nuintensitycor-3T" improved brain segmentation accuracy by optimizing the bias field correction [3.3.nuintensity].

Second, by invoking "-notal-check", we skipped the Talairach registration checks. Talairach registration was prone to failure especially in the infant subjects, and uneccessary for our further processing steps.

Third, we supplied and included T2-weighted MRI datasets with the options "-T2" and "-T2pial". The combination of T1- and T2-weighted images improves tissue differentiation especially around the pia mater, yielding a more accurate cortex segmentation. This pipeline yielded a continuous, antomically plausible cortical surface in MRI space.

Forward and inverse operator For the forward operator, three components were necessary: a source model, a BEM model and a coregistration file.

The cortical surface from Freesurfer was used to construct the source model. Sources were generated by the MNE function mne_setup_bem. The result were 20484 sources (10242 per hemisphere), distributed with approximately equal density over the cortical surface.

The head surface from Freesurfer was used to extract a scalp surface layer. The BEM was constructed from this scalp layer with the function mne_surf2bem, using the default options. This function sampled down the original surface to the 4th subdivision of an icosahedron. The finished BEM consisted of 5140 nodes.

Finally, a coregistration file provided the transformation between MRI space and MEG space. This coregistration attempted to minimize the distance between digitized head surface points and the head surface extracted from the MRI. It was performed for each subject individually using the software mne_analyze. The initial fit was done manually, with visual error feedback. The following fine adjustment was performed automatically. This process was repeated until the average spatial error was less than 2mm. These three components were assembled into a forward operator by the method mne_do_forward_solution.

For the inverse operator, three components were necessary again: the forward model, a noise covariance matrix, and a regularization factor. The forward model was supplied from the previous step. For the noise covariance matrix, the 1000ms after visual onset were extracted from each trial. Then, the covariance matrix was computed from this data with the function mne.compute_covariance(). The regularization factor was determined from this noise covariance matrix.

First, only coefficients from gradiometer channels were selected.

Second, these coefficients were transformed with a singular value decomposition.

Third, the upper cutoff was defined as the first value of the transformed coefficients.

Fourth, the index at which the transformed coefficients performed the steepest drop in logarithmic value was determined.

Fifth, this index was defined as the maximum amount of usable dimensions.

Sixth, the lower cutoff was defined as the value at this index, plus 15%.

Seventh, the regularization factor was computed by dividing the lower cutoff by the higher cutoff.

Each component was calculated individually for each subject. The three components were combined into the inverse operator by the method mne_do_inverse_operator. The regularization factor was supplied with the option "–megreg".

Inverse solution For determining regional cortical activity, 8 regions needed to be defined: the primary auditory cortex (PAC), the anterior and posterior parts of the superior temporal sulcus (a/pSTS), the anterior and posterior parts of the superior temporal gyrus (a/pSTG), Brodmann area 45 (BA45), Brodmann area 44 (BA44) and the ventral Brodmann area 6 (BA6v). The regions were spatially defined manually on the reference subject. Freesurfer provided the aparc.a2009s segmentation, which became the basis for this regional selection. Because of the insensitivity of MEG to perpendicular sources, a separation line between neighboring gyri and sulci needs to be completely unbiased, or it could lead to strong accidental misattribution. Because node resolution on lateral gyrus walls was not sufficient to ensure a bias-free separtion line, we combined the regions aSTS and aSTG into aSTS+G. Similarly, pSTS and pSTG were combined into pSTS+G. The final regions of interest on the reference brain are visualized in [Fig. 3.3.3.ROI]. Regions were automatically valid for all other subjects as well, since the inverse operator was already calculated in respect to the morphed reference brain.

[Fig. 3.3.ROI: Left: selected regions on the natural cortex. Right: selected regions on the inflated cortex.]

The inverse operator was used to calculate inverse solutions from MEG sensor data. Inverse solutions were calculated for each time point, region, trial and subject individually. The process was performed by the function mne.minimum_norm.apply_inverse_epochs(), with sLORETA as the inverse method. The option "pick_ori=normal" ensured that currents leaving and entering the cortex were designated positive and negative, respectively. Due to the combination of passive and active noise reduction and artifact suppression, we assumed a fairly high signal-to-noise-ratio (SNR) of 100:1 for each individual source. The regularization factor was estimated by $\frac{1}{SNR} = 10^{-4}$. The result was a series of activation patterns within each region. Finally, the mean of regional node activity was calculated for each time point, region, trial and subject.

The resulting localized activity was again subjected to a cluster analysis. Extracted trials were split into a 4 parts (2 groups x 2 conditions). The two groups were again evaluated separately. Trials contained average data from six regions (PAC, aSTS+G, pSTS+G, BA44, BA45 and BA6v). Clusters were again determined with the MNE function stats.permutation_cluster_test [8]. The function too was run with 2500 permutations, and an t-threshold of 1.0.

For visualization purposes, grand average activity was also calculated for each cortical region, group and condition.

3.3.4 Interaction analysis

The TRENtool software was used for exploring transfered entropy between cortical areas. - Role of embedding, Ragwitz optimization - statistical tests between conditions and multiple comparisons

Part II

Empirical Part

Chapter 4

Results

4.1 Behavioral results

If not noted otherwise, two comma-separated values in brackets describe the upper and lower values of a 95%confidence interval.

4.1.1 Response times

A Shapiro-Wilk test was conducted to determine if individual data were normal distributed. No subject exceeded a probability of p=1e-8 that their response times were normal distributed. Therefore, we decided to represent individual response times by their median.

Children needed a median time of $1.91s$ $(1.64s, 2.19s)$ to respond to object-relative clauses. For subject-relative clauses, they needed $1.97s$ $(1.69s, 2.25s)$. Adults needed a median time of $1.51s$ $(1.28, 1.73s)$ to respond to object-relative clauses. For subject-relative clauses, they needed $1.60s$ ($1.37s$, $1.82s$).

A Shapiro-Wilk test determined that response time data was normal distributed with a probability between 1.5% and 27%. A Levene's test determined that the probability of median response times being normal distributed was between

4.1.2 Response accuracy

For the analysis of variance (ANOVA), all data must be normal distributed with equal variance. A Shapiro-Wilk test determined that the probability of accuracy data being normal distributed was between 2.0 and 20.3%. A Levene's test yielded that the probability that accuracy data were distributed with equal variance was less than $p = 0.1\%$.

The ANOVA is known to be robust for considerable deviations from the normal distribution. However, it is highly vulnerable to violation the assumption of equal variances. To meet this requirement, we transformed the accuracy data with the inverse sigmoid function.

This procedure, however, created singularities in some extreme cases, i.e., when a subject performed with a 100% accuracy rate. To prevent this issue, we added a single incorrect trial to every subject's performance for the following analysis.

After the transformation, the same tests as before were conducted. The probability for transformed accuracy data being normal distributed was between 0.5% and 27%. The probability for transformed accuracy data being distributed with equal variance was $p = 72.2\%$. Supported by these findings, the transformed accuracy data was included in the ANOVA.

4.1.3 Analysis of combined performance data

The ANOVA was conducted with the transformed accuracy data and the median response time. Each subject provided one data point for each metric. Data were analyzed with a group x condition design. Accuracy estimates were transformed back with the sigmoid function $acc = 1/(1 + exp(-x))$.

Children responded 0.39s slower than adults ($1.94svs.1.55s$). This difference was significant ($F_56 = 9.4, p = 0.3\%$).

Children responded with an average accuracy of 93.8% (92.2%, 95.0%). Adults performed much better, with an average accuracy of 97.9% (97.5%, 98.3%). This difference was highly significant ($F_56 = 52$, $p = 1.6e-9$).

Sentence condition had no impact on median response times ($F = 0.33$, $p = 57\%$) or on response accuracy ($F = 1.3, p = 26\%$). There was no interaction effect between group and sentence condition ($F < 0.1, p > 80\%$).

4.2 Sensor–space activity

4.2.1 Preprocessing review

4.2.2 Comparison of regional activity

4.3 Reconstructed cortical activity

4.3.1 Inverse models review

4.3.2 Comparison of regional activity

Chapter 5

Discussion

5.1 Impact on hypotheses

5.2 Comparison to previous results

5.3 Limitations and possibilities of this study

[What do this data mean?] Sensor data: Confirmation by cluster analysis, and by blind comparison

Children: “Middle“: 183-350, positive, right parietal mags “Late“: 380-647, negative, left-temporal and left-frontal mags “Very Late“: 1455-1631, positive, (right-temporal grads) Adults: “Middle“: 282-429 (left-temporal grads) “Late“: 637-750 (left-parietal mags) “Very late“: 1013-1136 (right-temporal mags)

Source data:

[Are the hypotheses confirmed?]

[What have similar studies found out?]

[In which ways is my study limited, and what can it dare to say?]

Chapter 6

Conclusion

Appendix A

Appendix

References

Afifi, A. K., & Bergmann, R. A. (1998). *Functional neuroanatomy: Text and atlas*. McGraw-Hill: New York.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington: Winston.

Bekerian, D. A., & Baddeley, A. D. (1980). Saturation advertising and the repetition effect. *Journal of Verbal Learning and Verbal Behavior*, *19*, 17–25.

List of Figures

List of Tables

Curriculum Vitae

Name:	Max Mustermann
Geburtsdatum:	12.12.1978
Geburtsort:	Minsk

seit 2002	Doktorandin am Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig

Studium:

1999 – 2005	Universität
1998	Abi
1986 – 1998	Schule

Schulbildung:

1999 – 2005	Universität
	Universität
1998	Abi
1986 – 1998	Schule

Selbständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt wurde un d dass die aus fremden Quellen direkt oder indirekt übernommenen Gedanken in der Arbeit als solche kenntlich gemacht worden sind.

Dein Name
Leipzig, 11. Februar 2111