amsterdam brain & cognition

ournal

lssue 2 May 2015

IIS

0



××



Editorial Board of the ABC Journal						
Editorials	page 6					
Laurens Kemp Executive Control and Emotional Information Processing in Dissociation: A Task-Switching Study	page 9					
Reneé San Giorgi Hyperactivity in amygdala and auditory cortex in misophonia: preliminary results of a functional magnetic resonance imaging study	page 21					
Kim van Dijk Predicting acute trauma-related symptoms in recently traumatized individuals at increased risk for post-traumatic stress disorder - a tractography study	page 29					
<b>Rianne van Rooijen</b> Experience-dependent suppression of mu- and beta-power in the infant motor cortex while observing others' actions	page 41					
Willemijn van Woerkom Getting to the bottom of processing behinds	page 49					
Marieke Woensdregt Stats & Nets versus Rules & Symbols: Re-opening the Debate on Learning Mechanisms for Articial Grammar Learning	page 65					
Abstracts	page 81					
7 questions about academic publishing	page 85					

## Editorial Board of the ABC Journal



#### Editor in Chief Michael X Cohen

Founding Editors Cato Drion Steven Raaijmakers

Assistant Editors

Anna Lien Bouhuis David Iñaki López Mejía Myrthe Knetemann Siméon Lahaije Natalia Rivera Koen Seignette Esther Visser Josien Visser

#### **Research Master Brain and Cognitive Sciences Staff**

Programme Director Harm Krugers

Programme Manager Silke van Beekum

Assistant Programme Manager Tim Bulters

Study Advisors
Astrid Bosma



Dear reader,

The ABC Journal has developed impressively since its inception and first issue. What began as something of an experimental "start-up" is now growing into its own community, with new students joining the editorial team, new and streamlined submission/evaluation procedures, a constantly-updated Facebook page, and so on. Our monthly meetings have being increasingly dynamic and enthusiastic, and it's been a pleasure to be part of them.

It is now time for me to hand over the figurative torch to the next editor-in-chief. My future at the Radboud University beckons me, and I must part ways with the ABC Journal. The next editor-in-chief (to be named in the coming months) will find an excellent team of highly motivated students.

The ABC Journal showcases the best research masters project reports here at the UvA community. Enjoy this second issue of the ABC Journal, and congratulations to the students whose reports were selected to be printed in full and in abstract form.

Michael X Cohen Editor in Chief Amsterdam Brain and Cognition Journal



<u>abc</u>

Dear reader,

It has been a year since the first issue of the ABC Journal, and a lot has changed. The committee has welcomed three enthusiastic new members, Anna Lien, Esther and Josien, our Facebook page is liked and viewed by more and more people every week, new and experienced researchers are happy to answer our 7 questions about academic publishing, and this journal steadily becomes a more integrated part of the master Brain and Cognitive Sciences. Thanks to Silke, sending your internship article to us has now become part of the grading process, which means the ABC Journal feels more and more like a part of our master and we are very happy with this support.

For the current issue we have selected articles that cover a wide range of subjects, from misophonia to toddlers and from artifical grammar learning to monkey butts, and that also show the high quality of our master programme and the hard work of its students. Also included in this second issue are several abstracts of other research done in the past year, and 7 questions about academic publishing with three different researchers.

As the ABC Journal continues, we welcome new members, but also say goodbye to others. We want to thank Tineke and Natalia for their hard work and dedication, and Mike for setting up and leading the journal, and wish them all the best on their future endeavors.

For now, please enjoy the second issue of the ABC Journal.

Siméon Lahaije

Editor

Amsterdam Brain and Cognition Journal



The ABC Journal committee. Standing, I-r: Esther Visser, Siméon Lahaije, Anna Lien Bouhuis, Koen Seignette and Josien Visser. Sitting, I-r: Davíd López Mejía, Natalia Rivera, Michael X Cohen and Myrthe Kneternann.



# Executive Control and Emotional Information Processing in Dissociation: A Task-Switching Study

### Laurens Kemp

Supervisor: David Marcusson-Clavertz Center for Research on Consciousness and Anomalous Psychology, Department of Psychology, Lund University

#### ABSTRACT

Dissociation research has generated conflicting results on cognitive performance. We used a task-switching design to determine how dissociation relates to executive functions and emotional processing capabilities. Thirty-nine healthy participants performed an emotional switch task by categorizing faces based on sex or emotion and an affectively neutral control task. We found a relation between the switch cost for angry faces and the Depersonalization and Amnesia factors of the Dissociative Experiences Scale, showing an unexpected peak-shaped trend in which medium dissociatives had higher switch costs than lows and highs. The same relation appeared as a non-significant trend in the rest of the data. Aside from this, there was no general effect of dissociation on task performance. Contrary to our expectations, our results showed that the switch cost for angry faces was bigger when categorizing emotion than when categorizing sex. According to the theory of task set inertia, this might be explained by increased inhibition of salient emotional information carrying over across trials. Our study supports previous findings of increased sensitivity to emotional information for high dissociatives and suggests better performance by high and low than medium dissociatives.

#### **KEYWORDS**

Executive Control, Emotional Information Processing, Switch Task

#### INTRODUCTION

The construct of dissociation is complex and the definitions of the term vary in the phenomena they include or exclude. One of the senses of dissociation is "the coexistence of separate mental systems that should be integrated in the person's consciousness, memory, or identity" (Cardeña, 1994 p. 19, but see Braude, 2009). This definition is exemplified by dissociative amnesia, in which the person cannot access such bits of episodic information as one's own name. Another sense of dissociation is known as experiential detachment, including the experience of being outside oneself or as if one were only an observer of one's behavior (depersonalization), or the experience that the world is unreal and viewed as if through a foggy lens (derealization) . When dysfunctional and distressing, these phenomena are considered psychopathological symptoms, but dissociation has also been studied in non-pathological contexts. For instance, aspects of dissociation such as hypnotic automatism (e.g., the ability to act out post-hypnotic suggestions without being aware of them) and absorption (the ability to become so immersed into an activity that one is unaware of their surroundings) can be regarded as harmless or even beneficial to a person (Braude, 2009).

What separates normal from pathological dissociation remains unclear (Dalenberg & Paulson, 2009). One proposal is that dissociative experiences can be placed on a continuum: if they are fleeting or controlled, they do not have adverse effects, but chronic and uncontrolled dissociative experiences can lead to dysfunction (Braun, 1993). Another proposal is that there are some aspects of dissociation such as absorption that are, at least not to excess, not associated with psychopathology, whereas others such as spontaneous amnesia are more likely to reflect dysfunction (Waller, Putnam, & Carlson, 1996). This implies that there might not be a sharp cutoff where the level or frequency of dissociative experiences

#### EXECUTIVE CONTROL & EMOTIONAL INFORMATION

goes from benign to detrimental, and the cognitive abilities of dissociatives, such as working memory (WM) and emotional processing, may be differentially affected as the level of dissociation approaches the more extreme end of the continuum. Differences in cognitive performance between different levels of dissociation have been researched in a number of studies.

#### Working Memory

WM refers to the cognitive ability to temporarily store and manipulate a limited amount of information. WM performance, and how it relates to dissociation has been assessed in a variety of ways. De Ruiter, Phaf, Elzinga, and van Dyck, (2004) found enhanced verbal WM for high compared to low dissociatives, as measured by the Dissociation Questionnaire, (DIS-Q; Vanderlinden, Van Dyck, Vandereycken, & Vertommen, 1991). Amrhein, Hengmith, Maragkos, and Henning-Fast (2008) did not replicate this finding with a similar test using the Dissociative Experiences Scale (DES; Carlson & Putnam, 1993). Another study used a counting span task and found that participants with high DES scores performed worse than those with low scores (Terhune, Cardeña, & Lindgren, 2011). Elzinga et al. (2007) tested WM performance with an n-back paradigm and found that the performance of patients with a dissociative disorder did not deteriorate as much as that of healthy controls when WM load was increased. In an earlier review, Dorahy (2001) had concluded that patients with a dissociative disorder may also show WM deficits, but later results did not provide strong support for this assertion (Dorahy, McCusker, Loewenstein, Colbert, & Mulholland, 2006).

De Ruiter et al. (2004) proposed that their results on WM indicate that a moderate degree of dissociation results in heightened attentional capabilities. This idea is consistent with the notion that a person can dissociate the task demands from interfering external influences through absorption in the task. Absorption is one factor of a proposed three-factor solution for the DES, the other two being amnesia and depersonalization/derealization (Stockdale, Gridley, Ware Balogh, & Holtgraves, 2002). It has been proposed that absorption, while associated with a risk of developing a dissociative disorder, is not maladaptive of itself (Dell, 2009). Conversely, uncontrolled and chronic depersonalization and amnesia are basic components of the dissociative disorders (Spiegel & Cardeña, 1991). Recently, a two-factor solution-automatic processes and defensive mechanisms-was proposed and supported by a confirmatory factor analysis (Larøi, Billieuz, Defeldre, Ceschi, & Van der Linden, 2013; Olsen, Clapp, Parra, & Beck, 2013). The defensive mechanisms factor has a large degree of overlap with the depersonalization and amnesia factors, and is also associated with pathological dissociation (Larøi et al., 2013).

#### Emotional Processing

Research on how dissociation affects the processing of emotional information, such as the judgment of valence of emotional stimuli, is similarly equivocal. A study on free recall of emotional words indicated that high dissociatives remember fewer negative, trauma-related words (e.g., "rape," "incest," and "assault") compared to low

dissociatives, but this was not the case in a divided attention task (DePrince & Freyd, 1999). A reason may be that high dissociatives have impaired selective attention for emotional information, but other studies failed to support this explanation. Oathes and Ray (2008) found that high dissociatives were faster and more accurate in identifying the emotional valence of words and pictures than dissociatives. De Ruiter, Phaf, Veltman, Kok, and Van Dyck (2003) also found that high dissociatives responded faster to threat words than low dissociatives. In contrast, Dorahy, Middleton, and Irwin (2005) found slower responses to emotional than to neutral words in patients with dissociative disorders, although a later study was unable to replicate this finding (Dorahy et al., 2006).

The faster reaction time (RT) found by Oathes and Ray (2008) and de Ruiter et al. (2003) may be explained by a high attentional capability of high dissociatives based on their high absorption. However, emotional stimuli may still have an effect on information processing in dissociatives, as shown by DePrince and Freyd (1999). This is further supported by an fMRI study that showed diminished brain activation for emotional faces in patients with a dissociative disorder (Lemche et al., 2008).

#### Executive Functions

In a study using a large number of cognitive tests, a divided attention task showed that high dissociatives performed worse than low dissociatives (Amrhein et al., 2008). This was thought to be due to deficits in executive function, leading to a decreased ability to inhibit error responses. This theory was supported by the finding that high dissociatives show increased perseverance in random number generation (Giesbrecht, Merckelbach, Geraerts, & Smeets, 2004). An executive control deficit may have also played a role in the study by Terhune et al. (2011) due to the high demand for cognitive control in the counting span task. It is also possible that the faster responses to emotional stimuli by high dissociatives (Oathes & Ray, 2008) come with a drawback in other modalities, meaning that their responses may be more difficult to inhibit when the task demands it.

Overall, these studies show that people diagnosed with a dissociative disorder often perform worse on tasks that test emotional processing, WM, and executive control, while in nonclinical samples high scores on self-report measures of dissociation may be beneficial to performance if the task is not too complex. Further research into the executive functions of dissociatives is sparse, however, and the inconsistency of previous results may be clarified by more insight into how dissociation relates to executive functions.

A method of assessing executive function is to use taskswitching. This test loads WM by requiring the participant to keep track of multiple tasks at once and requires shifting attention to a changing stimulus dimension. Task-switching (shifting) performance correlated with inhibition and WM-updating in a study investigating latent variables behind measures of executive functions (Friedman et al., 2006), so it could be used to test for the effects of executive control and inhibition deficits in high versus low dissociatives.

#### Our Study

We decided to investigate how dissociation affects the capacity for executive function and emotional processing through a taskswitching paradigm that, to our knowledge, had not been used to study dissociation. In this paradigm, a person is required to classify a stimulus on different dimensions (e.g., the sex or the emotion of a face), leading to two different tasks. From one trial to the next, the dimension that needs to be responded to can either change or not. In trials in which the task changes (switch trials), RT tends to be longer than when the task stays the same (repeat trials), known as the switch cost. The use of faces with emotional expressions has a number of advantages. Compared to words, faces are more complex stimuli processed quickly and in a highly specialized way (Haxby, Hoffman, & Gobbini, 2000) and using faces in the study of dissociation can lend more ecological validity to the theory that dissociatives are more sensitive to emotional stimuli.

In order to explain the discrepancy between the performance of non-clinical high dissociatives and dissociative disorder patients, we hypothesized that there would be a curvilinear relation between the magnitude of dissociation and the effectiveness of executive functions, as well as of emotional processing capabilities, in our non-clinical sample. Our rationale was that a moderate level of dissociation would prove beneficial to tasks requiring executive functions because of the dissociatives' increased capacity for absorption, but as the level of dissociation approached pathological levels, we predicted that performance would start to decrease. Furthermore, we hypothesized that high dissociatives' increased sensitivity to emotional information. Higher levels of dissociation can also exacerbate this effect, leading to impaired executive functions when emotional information is concerned.

We expected that the potential increase in performance would be related to the participant's score on the absorption factor of the DES, while higher scores on the depersonalization and amnesia factors would be associated with decreased performance. We also evaluated the overall DES scores.

Because angry faces are considered more salient than happy faces (Fox et al., 2000; Hansen & Hansen, 1988), we expected trials with angry faces to have shorter RTs when goal-relevant and longer when irrelevant. We also expected that the order in which emotional stimuli were presented would affect RT depending on whether angry faces needed to be inhibited or not.

#### **METHODS**

#### Participants

Forty healthy participants, recruited from Lund and Malmö, Sweden, took part in the study and were compensated with a cinema ticket. Written consent was obtained prior to participation. Following testing, one participant's data were considered to be unreliable and were excluded, resulting in a sample of 39 participants (23 female; ages 18-36; Mage = 26;  $\Box$  age = 3.7). The participants were tested by the first author, a white male aged 23, fluent in English with

Dutch as a native language. He is a student at Master's level and has previously conducted two other behavioral studies with nonclinical samples. Participants gave written informed consent and the research followed the ethical guidelines from the Swedish Ethical Board for student theses.

#### Self-report measure

Participants completed the Swedish version (4 participants) and the English version (35 participants) of the Dissociative Experiences Scale (DES-II). The DES is a 28-item questionnaire in which the frequency of each item is rated on an 11-point scale ranging from 0% (never) to 100% (always; Carlson & Putnam, 1993; Swedish version: Körlin, Edman, & Nybäck, 2007). This version of the questionnaire asked the participants how often they have had these experiences in the last two weeks. In non-clinical samples, a score of 20 is usually used as a threshold to separate high from low dissociatives. The mean score of the 28 items on the DES has been shown to be a valid measure of non-pathological dissociation (Waller, Putnam & Carlson, 1996). Reliability was high for the scores on the Absorption factor ( $\alpha$  = .867) and the combined Depersonalization/Amnesia factors ( $\alpha$  = .773).

#### Computer Task

#### Stimulus Materials

The stimuli used in the task were 96 grey scale photos of front-facing faces of males and females with an angry or a happy expression (24 per sex-emotion combination). Each individual person's face occurred only once in the stimulus set. Extraneous features of the photos were removed by digitally overlaying a black oval frame over the photo. The photos had been rated on valence and level of arousal (see Lundqvist, Juth, & Öhman, 2014) in a previous test by participants from a different sample. In this test, the display duration and response window had no time limit but the participants were encouraged to respond quickly. The ratings showed no differences in arousal between the stimuli.

For the control task, two geometric figures (a circle or a triangle) of either red or blue color were used. Faces (400 x 544 pixels) and figures (150 x 150 pixels) were presented using E-Prime software (version 2.0.8.74; Psychology Software Tools, Sharpsburg, PA) running Windows XP Professional, SP3. The stimuli were presented against a white background in the center of an LCD computer display.

#### Task Description

Participants performed the two successive tasks using a taskswitching paradigm (Figure 1): the Face task, in which they switched between categorizing faces based on emotion or sex (adapted from de Vries & Geurts, 2012), and the Figure task, in which they switched between categorizing figures based on shape or color. The order of the tasks was counterbalanced. A practice session of 25 trials preceded each task, and each task ran for four blocks of 96 trials each, making a total of 384 trials. For both tasks, each trial started with a central fixation cross presented with a 200–300 ms jittered duration. The fixation cross was followed by a textual response cue lasting 300 ms, consisting of two words, one to the left and one to the right of the fixation cross. These indicated which feature of the stimulus the participant needed to respond to: Happy/ Angry (an Emotion Judgment trial) or Male/Female (a Sex Judgment trial) for the face task. The Figure task was analogous, but used color and shape. The left/right location of the cue (e.g., Happy/Angry versus Angry/Happy) was also counterbalanced between (but not within) participants. The cue was followed by a stimulus, to which the participant had 2000 ms to respond. A response was given by pressing one of two buttons on a response box with either the index finger or the middle finger of the right hand. A response was immediately followed by a new trial, or if there was no response within the 2000 ms response window, a new trial started automatically.

The distribution of the two conditions in each task was made so that one trial would be followed by another of the same subtask (a repeat trial) three times more often than it would be followed by a trial of the other subtask (a switch trial). For example, a Sex Judgment trial would be followed by another Sex Judgment trial three times more often than it would be followed by an Emotion trial. Thus, the switch-to-repeat ratio was 1:3. This distribution was chosen to maximize the effect of the switch cost (Monsell & Mizon, 2006). One exception was made to prevent too many consecutive repeat trials, so after five consecutive trials of the same condition a switch would follow. In all other ways, the trials were generated pseudo-randomly.

#### Procedure

At the start of the experiment session, participants were allowed to (re-)read the information letter, which stated that they would be doing a test on their cognitive flexibility and emotional processing capabilities. They then signed the informed consent form and completed the DES (4 participants had sent their DES earlier via e-mail). Once the questionnaire was completed, they were given verbal instructions on how to perform the task, with additional instructions during the practice trials if necessary. The experimenter remained masked to participants' dissociation scores throughout the session.

Practice trials were identical to normal trials with the exception that they provided the participant with feedback on their performance after each response. Feedback was shown on whether the response was correct, incorrect, or no response was given, the participant's reaction time on that trial, and the cumulative percentage of correct responses. Each participant was required to score over 80% correct on these trials before proceeding with the normal trials. If they scored lower than this, they were required to repeat the practice trials until a score of more than 80% was achieved. Participants were given a break between each block of the tasks. All participants were tested in a sound-insulated room with the experimenter seated on the opposite side of the room, facing away and able to monitor the participants' progress through a mirror. After the test was completed, participants were debriefed about the purpose of the study.

#### Analyses

Analyses were carried out with IBM SPSS Statistics 22, with significance set at < .05 (two-tailed). Analyses of sex differences were done with t tests, and any effect of age on the tasks was measured through correlation techniques. DES scores that fell outside three standard deviations of the mean were corrected to the next highest score, and participants with an error rate exceeding three standard deviations were excluded from the error analysis.

Analyses of task- and condition-specific effects were carried out with repeated-measures Analyses of Variance (ANOVAs). For a preliminary analysis of the effects of dissociation, DES scores were grouped according to their mean percentage scores: 0 to 10% for the Low group (n = 17, Mage = 25, Gage = 3.5), 10 to 20% for the Medium group (n = 12, Mage = 26,  $\sigma$ age = 3.0), and >20% for the High group (n = 10, Mage = 27,  $\sigma$ age = 4.9). These groups were used as a between-subjects factor in the ANOVAs. To test the main hypotheses, regression analyses were performed on the average RTs and number of errors. Trials that had an incorrect response, or followed a trial with an incorrect response, were excluded to eliminate the effect of post-error slowing. The first trial of each block was also excluded, as these could not be categorized as a switch or repeat trial. To investigate the effects of task, switch or repeat trials, and stimulus type, repeated-measures ANOVAs were performed on these factors. In order to preserve reliability, only averages of at least twenty trials per condition per participant were used in the initial analysis. Thus, initial tests for stimulus emotion and stimulus sex were included in separate factorial designs.

#### RESULTS

To test whether task performance was consistent across all four blocks, reliability analyses were performed for the mean switch cost per block for each task. Reliability was acceptable for the Face task ( $\alpha = .61$ ) and slightly lower for the Figure task ( $\alpha = .54$ ).

The performances of male and female participants on the average RTs and total number of errors across all conditions did not differ (all ps > .05). The participants' age did not correlate significantly with average RTs or the total number of errors (all ps > .05). Demographics were not analyzed further.

#### Face Task vs. Geometric Figure Task

#### RT

The RTs on the Face and Figure tasks were compared in a 2 (Task) × 2 (Switch) factorial design, showing a main effect of Task (F (1, 36) = 58.6, p < .001,  $\eta$ p2 = .620) and of Switch (F(1, 36) = 30.5; p < .001;  $\eta$ p2 = .459), indicating that the RTs were longer on the Face task than the Figure task (Figure 2). There was also an interaction between Task and Switch (F(1, 36) = 6.83; p = .013;  $\eta$ p2 = .159).

Separate tests of the Face and Figure tasks showed a significant switch cost effect that was stronger for the Figure task (Face task: F(1, 36) = 11.4; p = .002;  $\eta p2 = .240$ ; Figure task: F(1, 36) = 34.4; p < .001;  $\eta p2 = .488$ ). Other main and interaction effects were not significant, all ps > .05.



Figure 1. Trial sequence of the Face task. A. An example trial in which the participant has to respond to the emotion expressed by the person. This trial can be followed by either A or B. B. A trial following A, in which the participant has to respond to the emotion again, a repeat trial. C. A trial following A, in which the participant has to respond to the sex of the person, a switch trial.

#### Errors

One participant had a total error rate exceeding three standard deviations from the mean and was excluded from this analysis. A 2 (Task) × 2 (Switch) analysis showed a significant effect of Task (F(1, 35) = 23.2; p < .001;  $\Box$ p2 = .399), indicating more errors for the Face than the Figure task. Relative error rates were also higher for switch as compared with repeat trials (F(1, 35) = 14.2; p = .001;  $\Box$ p2 = .288). However, no significant interaction with dissociation was found (all ps > .05). In a one-way ANOVA, commission and omission error rates also did not differ between dissociation groups (all ps > .05).

#### Emotion and Sex Judgment Tasks

#### Happy vs. Angry Faces

Within the Face task, comparing the subtask, switch, and stimulus emotion factors resulted in complex interactions. RTs for the Emotion and Sex Judgment tasks for different stimulus types were compared in a 2 × 2 × 2 design with Task (emotion vs. sex), Switch (switch vs. repeat), and Emotion (happy vs. angry) as factors. There was a main effect of Task (F(1, 36) = 4.771; p = .036;  $\Box$ p2 = .117) and of Switch (F(1, 36) = 10.6; p = .002;  $\Box$ p2 = .227) and an interaction

between Task and Emotion (F(1, 36) = 5.80; p = .021;  $\Box$ p2 = .139) and Task, Switch and Emotion (F(1, 36) = 5.18; p = .029;  $\Box$ p2 = .126). Other effects were not significant, all ps > .05.

To further investigate these interactions, 2 (Switch) × 2 (Emotion) analyses were performed for the Sex and the Emotion Judgment tasks separately. In the Sex Judgment task, RTs for angry faces were shorter than for happy faces (F(1, 36) = 7.36; p = .010;  $\Box$ p2 = .170). For the Emotion Judgment task, there was a main effect of Switch (F(1, 36) = 12.5; p = .001;  $\Box$ p2 = .257), showing longer RTs for switch than repeat trials. There was also an interaction between Switch and Emotion (F(1, 36) = 7.54; p = .009;  $\Box$ p2 = .173). Analysis for only the happy faces on the Emotion Judgment task showed no significant differences (all ps > .05), but for the angry faces on the Emotion Judgment task there was a significant effect of Switch (F(1, 36) = 17.1; p < .001;  $\Box$ p2 = .322), indicating that switch trials had longer RTs for angry faces than happy faces (Figure 3).

#### Presentation Order of Happy and Angry Faces

The effects of stimulus emotion on RT were further investigated by grouping the trials according to whether the previous trial and the current trial showed a happy or an angry face. A 2 (Switch) × 4 (Emotion Order) analysis of the Emotion Judgment task showed a significant main effect of Switch (F(1, 36) = 11.7; p = .002;  $\eta p2 = .245$ ), indicating longer RTs for switch than repeat trials, and an interaction effect between Switch and Emotion Order (F(3, 34) = 2.98; p = .045;  $\eta p2 = .208$ ). Separate analyses for emotion order showed significantly longer RTs for switch than repeat trials for Angry→Angry (F(1, 36) = 17.0; p < .001;  $\eta p2 = .320$ ), Happy→Angry (F(1, 36) = 9.03; p = .005;  $\eta p2 = .201$ ) and Angry→Happy trials (F(1, 36) = 6.97; p = .012;  $\eta p2 = .162$ ), but not for Happy→Happy trials (p > .05; Figure 4).

Our analysis of the Sex Judgment task further illustrated an interaction between Switch and Emotion Order (F(3, 34) = 3.82; p = .018;  $\eta p2 = .253$ ), but no other effect was significant (ps > .05). Further tests showed longer RTs for switch than repeat trials for Angry→Angry (F(1, 36) = 6.88; p = .013;  $\eta p2 = .160$ ) and Happy→Happy (F(1, 36) = 5.31; p = .027;  $\eta p2 = .129$ ) trials, but the other conditions were not significant (ps > .05; Figure 5).

#### Male vs. Female Faces

In order to control for possible effects of male and female faces on the interaction between Switch and Emotion, a 2 (Switch) × 2 (Stimulus Sex) ANOVA was done on the trials with angry faces on the Emotion Judgment task. This showed that switch trials were slower than repeat trials (F(1, 36) = 18.3; p < .001;  $\eta$ p2 = .337), and that trials with female faces were slower than male faces (F(1, 36) = 9.39; p = .004;  $\eta$ p2 = .207). No other effect was significant, (ps > .05).

#### Presentation Order of Male and Female Faces

We used a 2 (Switch) by 2 (Sex Order) analysis to test for the effect of higher switch cost for consecutive trials that showed the same stimulus sex and found a main effect of Switch (F(1, 36) = 16.2; p < .001;  $\eta p2 = .310$ ), showing longer RTs to switch than repeat trials, but no other effects (all ps > .05).

#### Dissociation

In the previous repeated-measures ANOVAs, participants' grouped DES scored were used as a between-subjects factor, but this factor did not show any significant interactions with the other task conditions (all ps > .05). Dissociation scores were further analyzed as a continuous variable.

Curve estimation regressions of linear and quadratic functions were performed on RTs. The variables tested were the switch cost and the number of errors for the Face and Figure tasks, the switch costs for the Emotion and Sex Judgment tasks, and for the different stimulus emotions and sexes (Table 1). The predictor variables used were the means of the DES, the Absorption factor, and the combined Depersonalization/Amnesia factors (Stockdale et al., 2002). The correlation between the two factors was significant (R = 0,743, p < .001).

The ANOVA for the relation between the Depersonalization/ Amnesia factor and the switch cost for angry faces was significant for a linear and a quadratic function (Linear function: R = .387, F(1, 37) = 6.53, p = .015; Quadratic function: R = .526, F(2, 36) = 6.88, p = .003). The functions showed a downward trending linear function, and a peak-shaped quadratic function that was maximal between a score of 5% and 10%, indicating that switch costs for angry faces were lowest for high dissociatives and highest for medium dissociatives (Figure 6). No other analysis was significant (ps > .05).

#### DISCUSSION

This study investigated the effect of emotional stimuli and dissociation on the performance on a switch task. We hypothesized that RTs to angry faces would be shorter when task-relevant, and longer when task-irrelevant. We also hypothesized that dissociation and switch cost would show a curvilinear relation, and that the Absorption factor would be positively related to task performance, while the Depersonalization and Amnesia factors would be



Figure 2. Mean reaction times and mean switch costs in the Face task and the Figure tasks. Error bars indicate standard error of the mean (S.E.M.).

negatively related to task performance. Our results contradicted these hypotheses in most cases: RTs in trials with angry faces were slower than in trials with happy faces, and higher scores on the Depersonalization and Amnesia factors were associated with lower switch costs. However, our results provide tentative support for a curvilinear relation between dissociation and switch cost.

#### Face Task vs. Figure Task

Results showed that the Face task was more difficult than the Figure task, likely due to the added difficulty of recognizing emotion and sex in a person's face. This was reflected in the overall longer RTs, as well as in the higher number of errors, on the Face task compared to the Figure task. The switch cost was higher on the Figure than the Face task (see Figure 2), most likely because in the former, participants have a higher task-repetition benefit as the task is easier, leading to shorter RTs for repeat trials. Although we could not assess this directly, this explanation is consistent with the other results that indicate the Face task is more difficult.

However, it should be noted that in task-switching paradigms, switch costs are commonly in the order of hundreds of milliseconds (e.g., Paulitzki, Risko, Oakman & Stolz, 2008; de Vries & Geurts, 2012), while our results showed much smaller switch costs. This was in spite of our efforts to maximize the switch cost with a weighted 1:3 switch ratio (Monsell & Mizon, 2006) and a reduced response-to-stimulus interval (Monsell, 2003). It is possible that the reduced response-to-stimulus interval made the task too difficult and inflated the RTs on repeat trials so that the switch cost became smaller. This hypothesis is bolstered by our finding that the control task showed overall shorter RTs as well as a higher switch cost (see Figure 2). Nevertheless, the difference between switch and repeat trials was significant, indicating that the switch cost effect was reliable.

#### Emotion Judgment Task: Happy vs. Angry Faces

In the Emotion Judgment task, the switch cost was disproportionately large for angry faces (60 ms) compared to the other conditions (±20 ms) (see Figure 3). When participants were tasked with categorizing the emotion of a face after previously having attended to sex, seeing an angry face caused them to be substantially slower. It appears that there is a similar, yet smaller effect when an angry face is shown on a repeat trial in the Emotion Judgment task, but it was not significant. This directly contradicts our predictions, as we expected the (ir)relevance and salience of the stimulus emotion to be associated with faster or slower RTs, yet in this case the most relevant and salient stimulus showed the slowest RT.

This counterintuitive result may be explained by the theory of task set inertia, which posits that the cause of the switch cost is due to the competing influence of the previously activated task on the response of the current task (Allport, Styles, & Hsieh, 1994). A recent review by Vandierendonck, Liefooghe and Verbruggen (2010) supports this theory by demonstrating that an asymmetric switch cost occurs in task-switching studies in which one task is dominant over the other: switching from the dominant to the nondominant task (such as Stroop word-naming versus color-naming) results in a lower, not higher, switch cost than switching in reverse. This is because performing the non-dominant task requires the dominant response to be strongly inhibited, while performing the dominant task does not require as much inhibition of the nondominant response. This leads to a higher switch cost when the task switches from the non-dominant one to the dominant one, as the inhibition of the dominant one has to be overcome before a correct response can be made.

In our study it is not obvious that responding to emotion was more automatic than responding to sex (or vice-versa), yet the



Figure 3. Mean difference between reaction times on switch and repeat trials for task conditions and stimulus emotions. Error bars indicate S.E.M. Bottom groups indicate task (categorizing emotion or sex); top groups indicate stimulus emotion (angry or happy faces).

#### **EXECUTIVE CONTROL & EMOTIONAL INFORMATION**

increased switch cost in response to angry faces on trials that switched from sex to emotion shows that the situation might be similar. We predicted that emotional information of negative valence (angry faces) must be inhibited in order to perform the Sex Judgment task, causing longer RTs compared to happy faces, but the reverse seems to be true: Inhibition causes RTs to be longer on the Emotion Judgment task when angry faces are presented, as the inhibition of emotional information must be overcome when a sex trial switches to an emotion trial. This causes RTs on these trials to be longer than trials in which two consecutive happy faces are presented, as those are not inhibited as strongly.

#### Sex Judgment Task: Happy vs. Angry Faces

When looking at the RTs for the Sex Judgment task, however, this explanation does not seem to suffice. RTs to angry faces were shorter than to happy faces, but when trials were split according to the emotion of the preceding stimulus, the results showed that an interaction effect underlay this result. There are significant switch costs when the previous and current stimulus emotions are the same, but no switch cost when they are different (see Figure 5). It could be that when two stimuli with the same emotion are presented consecutively and the task switches, there is a stimulus-response association that causes a stronger tendency to repeat the same response as on the previous trial, leading to a higher switch cost. If the stimulus emotion changed from the previous trial, the stimulusresponse association is not present, and switching to the other task is easier.

If this explanation is correct, we should have seen the same pattern in the Emotion Judgment task, with lower switch costs when the stimulus sex changed from one trial to another than when it stayed the same. Although the means indicated that the switch costs were larger when two consecutive stimuli had the same sex than when they were different, the difference was not significant. This may indicate the asymmetry of the task, as we can expect the more



Figure 4. Mean difference between reaction time on switch and repeat trials, and the presentation order of stimulus emotion, in trials where participants had to categorize emotion. Error bars indicate S.E.M.



Figure 5. Mean reaction times between switch and repeat trials, and the presentation order of stimulus emotion, in trials where participants had to categorize sex. Error bars indicate S.E.M.

#### KEMP

salient emotional information to have a stronger effect on response habits. This explanation is also consistent with the theory of task-set inertia, as it predicts a multitude of different interference effects that may influence switch costs (Vandierendonck et al., 2010). Although previous studies have not reported differential switch costs when a stimulus dimension repeats and the task switches, it is consistent with the notion of past trials affecting current performance that is at the basis of the task set inertia theory.

Although these data are consistent with previous results on taskswitching, there is a caveat: in order to test the effect of the stimulus emotion on the previous trial, the RTs for the switch trials had to be made of an average of ten trials per participant per condition, whereas previous tests had at least twice that. This means the within-subject variability in this test was higher and the measures not as reliable.

In the analyses of dissociation scores as a continuous variable, for the Amnesia/Depersonalization and the Defensive Mechanisms factors, the relation between the DES score and the RT on angry faces was significant, showing a downward trend for higher DES scores (see Figure 6). Although the distribution of the dissociation scores was skewed towards lower scores, the quadratic function indicates that the level of dissociation was associated with an increase in the switch cost for angry faces at intermediate scores, and a decrease at high scores. In fact, the seven participants who scored over 13% on the Amnesia/Depersonalization factors had switch costs that were close to zero and even negative, so they did not slow down on task-switching when angry faces were presented. This is in spite of our finding that in many conditions, angry faces were associated with significant switch costs.

These results show that high dissociatives are faster to respond to emotion of negative valence specifically, which is partially consistent with previous findings that high dissociatives are faster at categorizing emotional valence. Contrary to our hypothesis, the Amnesia and Depersonalization factors underlay this effect, not the Absorption factor. This is surprising as these factors are more associated with pathological forms of dissociation than absorption (Ross, Ellason & Anderson, 1995; Stockdale et al., 2002). Given that pathological dissociation is associated with cognitive deficits (e.g., Dorahy, 2001) one would not expect that individuals scoring higher on the factors deemed pathological would perform better on task-switching. On the other hand, because absorption has been argued not to be a part of dissociation (Dell, 2009), the other two factors may better represent other forms of dissociation—such as experiential detachment and dissociative amnesia—that have a bigger impact on cognitive performance. It should also be noted that factor solutions of the DES are in dispute and other factor solutions other than the one used here may prove to be more suitable in the future. Furthermore, the correlation between the factors was very high, leaving only a small amount of additional variance to be explained by separate analyses.

The smaller switch cost in response to angry faces for the high dissociatives, can be interpreted as emotional information of negative valence being processed more effectively by high dissociatives (De Ruiter et al., 2003). Alternatively, since the magnitude of the switch cost for angry faces was shown to be partially due to the inhibition of faces with negative valence, it is possible that the reduced switch cost was caused by an inhibition deficit in high dissociatives. This would paradoxically result in faster RTs to angry faces, as low dissociatives actually have longer RTs due to inhibiting the relevant stimuli. This would be further supported by the hypothesized effect of an inhibition deficit for irrelevant stimuli, yet this was not among our results. However, a reduced switch cost caused by an inhibition deficit would be consistent with our hypothesis, although the current results provide only weak evidence for this.

Aside from the relation between the switch cost for angry faces and dissociation, we did not find other significant effects when analyzing the continuous dissociation scores. However, a pattern was visible in the curve estimations for the other switch costs, namely that a quadratic function consistently showed a better fit than a linear function, and that all quadratic functions were similarly peakshaped. If this effect persists in a larger sample, it would indicate a performance deficit for medium dissociatives. As previous studies have investigated dissociation by only testing participants with very low or very high DES scores (e.g. Oathes & Ray, 2008; de Ruiter et al., 2004), the performance of medium dissociatives may have gone unexamined. It is worth noting is that this relation is the reverse of our hypothesis. Although this result is tentative, it warrants further investigation with the whole spectrum of dissociation.

Table 1. Means and standard deviations for the test variables used in the regression analysis. SC = Swith cost.

	Face task: SC	Face task: # of	Figure task: SC	Figure task: # of	Emotion Judgmen t task: SC	Sex Judgme nt task:	Happy faces: SC	Angry faces: SC	Female faces: SC	Male faces: SC
Variable		errors		errors		SC				
Mean (Std.	30,7	35,0	54,7	23,0	41,7	19,9	20,9	39,6	34,1	28,7
Deviation)	(53,7)	(16,1)	(58,8	) (19,5)	(66,7)	(70,9	) (57,9)	(65,1)	(60,7)	(64,4)



Figure 6. Curve estimation of a linear and quadratic function for the switch cost on angry faces and the mean score on the Amnesia and Depersonalization factors on the DES.

#### CONCLUSIONS

We found a multitude of condition- and stimulus-specific effects that are well-explained by the theory of task set inertia (Allport et al., 1994), and these show that inhibition and negatively-valenced emotional information play an important role in the task. As these features are known to be affected by dissociation (Amrhein et al., 2008; Giesbrecht et al., 2004), a switch-task with emotional faces should be informative in investigating dissociation.

We did not find that the level of dissociation had a strong effect on switch task performance in general. Although high dissociatives appear to be more sensitive to emotional faces, there was no strong indication that high levels of dissociation were associated with deficits in executive control or response inhibition. Nevertheless, the results on the different factors of the DES show that absorption and depersonalization are more closely related to changes in task performance than absorption.

To our knowledge, this is the first application of the switch task with emotional faces to the study of dissociation and it shows potential for use in future research. Further studies can investigate the effect of inhibition and stimulus-response associations and how they are affected by dissociation. The distribution of scores over selfreport scales of dissociation also warrants further investigation over a broader range of scores. This should provide further insight into the nature of performance benefits and deficits in non-pathological dissociation.

#### REFERENCES

- Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), Attention and performance XV: Conscious and nonconscious information processing (pp. 421–452). Cambridge, MA: MIT Press.
- Amrhein, C., Hengmith, S., Maragkos, M., & Hennig-Fast, K. (2008). Neuropsychological characteristics of highly dissociative healthy individuals. Journal of Trauma & Dissociation, 9, 525–542.
- Braude, S. E. (2009). The conceptual unity of dissociation: A philosophical argument. In P. F. Dell & J. A. O'Neil (Eds.), Dissociation and the dissociative disorders: DSM-V and beyond (pp. 27–36). New York, NY: Routledge.
- Braun, B. G. (1993). Multiple personality disorder and posttraumatic stress disorder. In J. P. Wilson & B. Raphael (Eds.) International handbook of traumatic stress syndromes (pp. 35–47). New York, NY: Plenum.
- Cardeña, E. (1994). The domain of dissociation. In S. J. Lynn & J. W. Rhue (Eds.), Dissociation: Clinical and theoretical perspectives (pp. 15-31). New York, NY: Guilford.
- Cardeña, E. (2008). Dissociative disorders measures. In A. J. Rush Jr., M. B. First, & D. Blacker (Eds.), (2008). Handbook of psychiatric measures (pp. 587-599). Arlington, VA: American Psychiatric Publishing.

KEMP

- Cardeña, E. & Carlson, E. (2011). Acute Stress Disorder revisited. Annual Review of Clinical Psychology, 7(1), 245-267.
- Carlson, E. B., & Putnam, F. W. (1993). An update on the Dissociative Experiences Scale. Dissociation, 6, 16–27.
- Dalenberg, C. J., & Paulson, K. (2009). The case for the study of "normal" dissociation processes. In P. F. Dell, & J. A. O'Neil (Eds.), (1994). Dissociation and the dissociative disorders: DSM-V and beyond (pp. 145-154). New York, NY: Routledge.
- Dell, P. F. (2009). Understanding dissociation. In P. F. Dell & J. A. O'Neil, (Eds.) (1994). Dissociation and the dissociative disorders: DSM-V and beyond (pp. 145-154). New York, NY: Routledge.
- Dorahy, M. J. (2001). Dissociative identity disorder and memory dysfunction: The current state of experimental research and its future directions. Clinical Psychology Review, 21, 771-795.
- Elzinga, B. M., Ardon, A. M., Heijnis, M. K., De Ruiter, M. B., Van Dyck, R., & Veltman, D. J. (2007). Neural correlates of enhanced working-memory performance in dissociative disorder: A functional MRI study. Psychological Medicine, 37, 235-245.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. Psychological Science, 17, 172-179.
- Giesbrecht, T., Merckelbach, H., Geraerts, E., & Smeets, E. (2004). Dissociation in undergraduate students: Disruptions in executive functioning. Journal of Nervous and Mental Disease, 192, 567-569
- Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. Journal of Personality and Social Psychology, 54, 917-924.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. Trends in Cognitive Sciences, 4, 223-233.
- Körlin, D., Edman, G., & Nybäck, H. (2007). Reliability and validity of a Swedish version of the Dissociative Experiences Scale (DES-II). Nordic Journal of Psychiatry, 61, 126-142.
- Larøi, F., Billieux, J., Defeldre, A. C., Ceschi, G., & Van der Linden, M. (2013). Factorial structure and psychometric properties of the French adaptation of the Dissociative Experiences Scale (DES) in non-clinical participants. Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology, 63, 203-208.
- Lemche, E., Anilkumar, A., Giampietro, V. P., Brammer, M. J., Surguladze, S. A., Lawrence, N. S., ... & Phillips, M. L. (2008). Cerebral and autonomic responses to emotional facial expressions in depersonalisation disorder. British Journal of Psychiatry, 193, 222-228.
- Levin, R., Sirof, B., Simeon, D., & Guralnick, O. (2004). Role of fantasy proneness, imaginative involvement, and psychological absorption in depersonalization disorder. Journal of Nervous

and Mental Disease, 192, 69-71.

- Lundqvist, D., Juth, P., & Öhman, A. (2013). Using facial emotional stimuli in visual search experiments: The arousal factor explains contradictory results. Cognition & Emotion, 26, 1012-1029.
- Oathes, D. J., & Ray, W. J. (2008). Dissociative tendencies and facilitated emotional processing. Emotion, 8, 653-661.
- Olsen, S. A., Clapp, J. D., Parra, G. R., & Beck, J. G. (2013). Factor structure of the Dissociative Experiences Scale: An examination across sexual assault status. Journal of Psychopathology and Behavioral Assessment, 35, 394-403.
- de Ruiter, M. B., Phaf, R. H., Veltman, D. J., Kok, A., & van Dyck, R. (2003). Attention as a characteristic of nonclinical dissociation: An event-related potential study. Neuroimage, 19, 376-390.
- de Ruiter, M. B., Phaf, R. H., Elzinga, B. M., & van Dyck, R. (2004). Dissociative style and individual differences in verbal working memory span. Consciousness and Cognition, 13, 821-828.
- Ross, C. A., Ellason, J. W., & Anderson, G. (1995). A factor analysis of the Dissociative Experiences Scale (DES) in dissociative identity disorder. Dissociation, 8, 229-235.
- Spiegel, D., & Cardeña, E. (1991). Disintegrated experience: The dissociative disorders revisited. Journal of Abnormal Psychology, 100, 366-378.
- Stockdale, G. D., Gridley, B. E., Balogh, D. W., & Holtgraves, T. (2002). Confirmatory factor analysis of single and multiple-factor competing models of the dissociative experiences scale in a nonclinical sample. Assessment, 9, 94-106.
- Terhune, D. B., Cardeña, E., & Lindgren, M. (2011). Dissociative tendencies and individual differences in high hypnotic suggestibility. Cognitive Neuropsychiatry, 16, 113-135.
- Vanderlinden, J., Van Dyck, R., Vandereycken, W., & Vertommen, H. (1991). Dissociative experiences in the general population in Belgium and The Netherlands: An exploratory study with the Dissociation Questionnaire (DIS-Q). Dissociation, 4, 180–184.
- de Vries, M., & Geurts, H. M. (2012). Cognitive flexibility in ASD; Task switching with emotional faces. Journal of Autism and Developmental Disorders, 42, 2558-2568.
- Waller, N., Putnam, F. W., & Carlson, E. B. (1996). Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. Psychological Methods, 1, 300-321.
- Watson, D. (2003). Investigating the construct validity of the dissociative taxon: Stability analyses of normal and pathological dissociation. Journal of Abnormal Psychology, 112, 298-305.



# Hyperactivity in amygdala and auditory cortex in misophonia: preliminary results of a functional magnetic resonance imaging study

### Renée San Giorgi

Supervisor: Arjan Schröder

Academic Medical Center, University of Amsterdam

#### ABSTRACT

Misophonia is a relatively unexplored condition where people react with anger, disgust, and rage when exposed to specific sounds. Previous studies have given indication that neurobiological abnormalities could possibly underlie this phenomenon. To investigate which brain regions are related to misophonia, we studied the BOLD response in reaction to misophonia trigger sounds using fMRI. Both misophonia patients and healthy controls were exposed to a blocked audio-visual stimulus provocation paradigm with 3 conditions (misophonia, aversive, neutral). Results indicated that patients showed increased activity in bilateral superior temporal cortex during misophonia condition, compared to the neutral condition. Furthermore, the left amygdala of patients was hyperactivated in the misophonia condition, compared to the aversive condition. These results highlight the role of patients' increased vigilance during exposure to misophonia trigger sounds. These findings are the first direct indication of abnormalities in brain function in misophonia and further research is suggested.

#### **KEYWORDS**

Misophonia, fMRI, Amygdala, Auditory Cortex

#### INTRODUCTION

Misophonia is a relatively unexplored condition where people experience hatred (miso) of certain specific sounds (phonia). It is characterized by reactions of anger, anxiety, disgust, and rage when exposed to specific, often human related, sounds (Edelstein, Brang, Rouw, & Ramachandran, 2013; Hadjipavlou & Baer, 2008; Kandel, Schwartz, & Jessell, 2000; Schröder, Vulink, & Denys, 2013). Every misophonia patient has their own set of trigger sounds, which typically includes oral and nasal sounds such as eating and breathing sounds. However, it can also consist of other sounds, such as finger tapping, typing, footsteps, clock ticking and many other diverse sounds. Immediately following the trigger sound, an autonomous, aversive physical reaction arises (Edelstein et al., 2013; Schröder et al., 2013). Misophonia patients report the impulsive urge to act aggressively by screaming at or attacking the source in order to make it stop (Schröder et al., 2013). This reaction disrupts the social functioning of misophonia patients and can have devastating effects

on their personal and professional lives. Much is still unknown about this extreme aversive reaction following relatively harmless sounds. More knowledge will benefit the development of effective therapy, and possibly give insight to the process of aggression in general.

Currently, misophonia cannot be classified in the psychiatric classification systems DSM-IV-TR, DSM-5 or the ICD-10. There are resemblances with several disorders that are mentioned in these classification systems, but it is not possible to attribute its distinct symptom pattern as a whole to any of these disorders (Schröder et al., 2013). Furthermore, misophonia does not seem to be related to a general hearing impairment (Edelstein et al., 2013; Schröder et al., 2014, 2013). Although comorbidity with tinnitus has been reported (Edelstein et al., 2013; Jastreboff & Jastreboff, 2006; Sztuka, Pospiech, Gawron, & Dudek, 2010), a significant distinction between tinnitus and misophonia can be made: tinnitus patients suffer from a phantom auditory perception, but misophonia is

different from hyperacusis in that it is not related to the physiological characteristics of the sound and that it does not lead to a startle reaction (Ferreira, Harrison, & Fontenelle, 2013).

Although scientific interest in misophonia has grown over the past two years, it yet needs to be thoroughly investigated on a larger scale. Several case studies have been published (Ferreira et al., 2013; Kluckow, Telfer, & Abraham, 2014; Neal & Cavanna, 2013; Webber, Johnson, & Storch, 2013) but few studies have included larger groups of patients with misophonia symptoms. A recent study proposed diagnostic criteria for misophonia (Schröder et al., 2013), and thus set the first step in classifying this condition as a psychiatric disorder. In another study, higher skin conductance responses (SCRs) to auditory stimuli were found in misophonia patients (Edelstein et al., 2013). This contributes to the idea that a misophonia reaction is impulsive, autonomous, involuntary, and initially not a cognitive process.

The current study aims to clarify which brain areas are involved in a misophonia reaction using functional magnetic resonance imaging (fMRI), by implementing an audio-visual stimulus provocation paradigm. If abnormalities are found, it would further strengthen the idea that misophonia should be classified as a distinct psychiatric disorder. It would also take the first step in further research on specific aspects of this difference. For example, further experiments can focus on the effect of cognitive behavioural therapy on brain functioning, the stimulus properties that are crucial for a misophonia reaction, or the causes for these deviant brain responses in misophonia patients.

As of yet, the only research currently known to have investigated the neurobiological properties of misophonia, described an electroencephalography (EEG) experiment (Schröder et al., 2014). Using an auditory oddball task, the experimenters identified a diminished N1 component. This indicates an underlying neurobiological deficit in automatic auditory processing in misophonia patients, which calls for more investigation.

In order to properly study the aversive nature of a typical misophonia reaction, two research questions are addressed in the current study. (1) Which brain areas show different blood-oxygen level-dependent (BOLD) responses in misophonia patients compared to healthy controls (HCs), when exposed to stimuli triggering a misophonia reaction, compared to neutral stimuli? (2) Which brain areas show different BOLD responses in misophonia patients compared to HCs, when exposed to stimuli triggering a misophonia reaction, compared to aversive stimuli?

To test this, participants are scanned while exposed to a blocked audio-visual stimulus provocation design consisting of three categories: misophonia (video clips with trigger sounds), aversive (violent or disgusting video clips), and neutral (silent video clips that do not evoke a strong emotional reaction). The reasoning behind this is that misophonia patients are prone to experiencing the misophonia video clips as aversive, while healthy controls will more likely experience them similar to neutral video clips.

Even with the lack of previous fMRI experiments with regard to misophonia, certain predictions for results in this study can be made. Several brain areas should be under close consideration based on existing literature on other fMRI experiments. These regions of interest (ROIs) can be separated into two categories. First, areas related to affect are expected to respond differently in misophonia patients during stimulus provocation. This includes areas related to aggression and aversion, such as the amygdala (Nitschke, Sarinopoulos, Mackiewicz, Schaefer, & Davidson, 2006). The amygdala is related to negative affect and associative aversive learning (Davidson & Irwin, 1999), and the importance of the amygdala in attentional functioning and vigilance has been stressed in several studies (Davis & Whalen, 2001; Holland & Gallagher, 1999). Since misophonia patients are highly focused on their trigger sounds, we hypothesize that the amygdala is hyperactivated in misophonia patients.

Second, areas related to auditory processing are ROIs. These areas are located in the superior temporal cortex and include the primary (A1) and secondary (A2) auditory cortex, as well as the surrounding association cortex. One study suggests that both A1 and A2 are important for selective auditory attention (Jäncke, Mirzazade, & Shah, 1999) and another study highlights the importance of the auditory association cortex in modulation of auditory attention (Grady et al., 1997). Since misophonia patients have an increased focus on their trigger sounds, we hypothesize hyperactivation in these areas. This report will only describe the preliminary results because data acquisition is currently still underway.

#### **METHODS**

#### Participants

Ten patients with misophonia (7 females, aged 18-44 years, mean = 32.2 years, SD = 8.7 years) who matched the diagnostic criteria for misophonia as proposed by Schröder et al. were recruited. All patients scored the Amsterdam Misophonia Scale (AMisoS) (Schröder et al., 2013) (score 10-18 points, mean = 14.5, SD = 2.3), which indicated that patients showed either moderate or severe symptoms of misophonia. 7 HCs (4 females, aged 21-51 years, mean = 29.4, SD = 11.1) matched for age and gender and with no misophonia symptoms or psychiatric disorders were scanned as well.

Furthermore, patients were excluded if they experienced less than 3 out of the 4 stimuli used in our experiment (carrot eating, grapefruit slurping, heavy breathing, and typing) as a trigger sound. This was assessed using the Misophonia Sound List (Misofonie Geluidenlijst) where patients score typical misophonia trigger sounds on a 5-level Likert scale. One patient reported experiencing only 2 out of 4 stimuli as trigger stimuli, but was not excluded.

The misophonia patients followed no or minimal misophoniaspecific treatment prior to testing (with a maximum of 2 misophonia group therapy sessions), with the exception of 1 participant who followed full group therapy. However, this participant still showed sufficient misophonia symptoms in order to be included in this

#### SAN GIORGI

study. Presence of major depression, major anxiety disorder, bipolar disorder, autism spectrum disorders, schizophrenia or any other psychotic disorder, as well as substance related disorders during the past 6 months were exclusion criteria. Patients who were taking benzodiazepines, antidepressants or stimulants were excluded. We excluded HCs when they reported misophonia complaints or suffered from any psychiatric disorder or significant other medical conditions.

Participants from both groups were between age 18 and 65, did not suffer from hearing damage, had no history of head trauma, and were screened for having no MRI contraindications. The study was approved by the Medical Ethics Trial Committee (METC) of the Academic Medical Center (AMC) Amsterdam and written informed consent was obtained for all participants.

#### Experimental design

During stimulus provocation, participants were exposed to audiovisual stimuli while lying inside the MRI scanner. The video clips consisted of three categories: misophonia, aversive, and neutral. The stimulus provocation paradigm was constructed as a blocked design, with video clips for the 3 conditions (misophonia, aversive, neutral) and 1 fixation period, all with 25 second duration and with a 2 second pause between clips. There were 4 video clips per condition, resulting in a total duration of (27 s x 4 x 4 =) 7 minutes and 12 seconds. The order of conditions was fixed in a pseudorandom order, but the sequence of the video clips within a condition was randomized (See Fig. 1).

The misophonia condition consisted of video clips where typical misophonia trigger sounds (carrot eating, grapefruit slurping, heavy breathing, and typing) were produced by a male actor. The aversive condition consisted of violent or disgusting video clips obtained from various commercially available movies (Boyle, 1996; Campion, 1993; Kaye, 1998; Noë, 2002). The neutral condition consisted of video clips depicting a male actor performing soundless activities, such as

reading and meditating. A previous pilot study outside the scanner confirmed that the aversive video clips evoked aversive emotions in both misophonia patients (n = 2) and HCs (n = 6), while misophonia video clips were considered aversive by misophonia patients only, and neutral by HCs. The neutral videos were experienced as neutral by both groups.

#### MRI data acquisition

Magnetic resonance images were obtained using a Philips Ingenia 3.0T MRI system (Philips Medical Systems, Best, the Netherlands), equipped with a SENSE 32 elements head coil. During stimulus presentation, T2\*-weighted BOLD images were acquired using Echo Planar Imaging (EPI). Per volume 37 transversal slices were acquired (in-plane resolution  $3 \times 3$  mm, slice thickness = 3 mm; slice gap = 0.3 mm; TR/TE = 2000/27 ms, 80 x 80 matrix). A T1-weighed structural image was obtained for spatial normalization during fMRI pre-processing (3D MP-RAGE). 180 sagittal slices were acquired (voxel size = 1 mm3, TR/TE = 7000/3.2 ms, 256 x 256 matrix, FOV = 256 x 240 mm).

#### MRI data analysis

MRI data analysis was performed with SPM8. Images were realigned in order to correct for head movement, corrected for slice time acquisition, and coregistered to the MP-RAGE anatomical scan. Next, they were normalized to MNI template, resampled to 2x2x2 mm, and spatially smoothed (8 mm FWHM).

The effect of the three different video clip conditions was estimated. These conditions were modelled as boxcar regressors, and convolved with a hemodynamic response function (HRF), as provided in SPM8. Also, the realignment parameters were added as regressors. Then, a high-pass filter (1/128 Hz) was added and serial correlation was accounted for using an autoregressive model (AR(1)). Contrast images comparing the neutral condition with both the aversive and misophonia condition across groups were acquired,



Figure 1. Paradigm for stimulus provocation. Each video clip lasted 25 s, with 2 s breaks in between. The order of conditions was fixed in a pseudorandom order, but the order of video clips within a condition was randomized. For example, all trials started with a neutral video clip, but in each trial the neutral video clip had been randomly chosen out of all 4 neutral video clips.

#### HYPERACTIVITY IN AMYGDALA AND AUDITORY CORTEX IN MISOPHONIA

in order to verify that this paradigm does evoke different brain responses in the misophonia and aversive condition compared to the neutral condition. Then, group differences were first investigated between the misophonia and neutral condition, and subsequently between the misophonia and aversive condition. Results were further investigated with ROI analysis, using the bilateral amygdala and the bilateral superior temporal cortex. Finally, if ROI analysis provided significant results, simple effect tests were performed to track down the underlying origin of this significant result. In order to correct for multiple comparisons, statistical tests were family-wise corrected (FWE) for the whole brain or small volume corrected for the ROI, if applicable. The ROIs were defined using the Anatomical Automatic Labeling (AAL) toolbox for SPM (Tzourio-Mazoyer et al., 2002).

#### RESULTS

## Effect of non-neutral versus neutral condition across groups

Initially, in order to investigate if the non-neutral conditions activate the ROIs in general, BOLD responses of all participants during exposure to misophonia and aversive video clips were compared to activity during exposure to neutral video clips. After whole brain FWE-correction, significant results were found in visual cortex, auditory cortex, and affective brain areas such as the amygdala and the anterior insula (see Table 1, see Fig. 2).

### Effect of misophonia versus neutral condition between groups

ROI analysis consisting of the superior temporal cortex revealed a significant interaction effect (p = 0.04). The right auditory cortex (x = 54, y = -26, z = 2) was significantly more active in misophonia patients compared to HCs, when comparing the misophonia condition with the neutral condition (see Fig. 3; see Table 2 for an overview of all group effects).

A simple effect test using the superior temporal cortex as ROI to compare BOLD response from patients during the misophonia condition with BOLD response from HCs during the misophonia

condition reveals no significant effect. However, with the superior temporal cortex as ROI, a significant effect was found when comparing BOLD response during the misophonia condition with BOLD response during neutral condition in patients only. These effects were located in the left (p = 0.00, [x = -54, y = -34, z = 10]) and right (p = 0.00, [x = 54 -26 2]) auditory cortex (see Fig. 4). The same hyperactivation of the superior temporal cortex was found for HCs (left: p = 0.00, [x = -56, y = -8, z = 6]; right: p = 0.00, [x = 66, y = -12, z = 0) (see Fig. 5).

After whole brain FWE-correction, no significant results were found. ROI analysis consisting of the amygdala showed no significant effects.

## Effect of misophonia versus aversive condition between groups

ROI analysis of the amygdala revealed a significant interaction effect (p = 0.05) in the left amygdala (x = -16, y = -2, z = -12), where misophonia patients displayed greater activity than HCs, when comparing the misophonia condition with the aversive condition.

Using the amygdala as ROI revealed no significant effect in misophonia patients when comparing the aversive condition with the misophonia condition. However, HCs do show a significant effect in the left amygdala (p = 0.01, [x = -18, y = -4, z = 14]) when comparing their aversive condition with their misophonia condition. Further simple effect testing revealed that the left amygdala (x = -16, y = -2, z = -14) is significantly more active in HCs (p = 0.02) than in misophonia patients during the aversive condition.

After FWE-correction for the whole brain, no significant results were found. ROI analysis using the superior temporal cortex showed no significant effects.

#### DISCUSSION

These preliminary results suggest the presence of several functional abnormalities in brain response of misophonia patients during their experience of misophonia triggers. Compared to neutral stimuli, misophonia patients showed hyperactivation in the bilateral auditory cortex when exposed to misophonia triggers. In addition, the left

Table 1. Brain regions with larger responses during non-neutral (misophonia and aversive) conditions than during the neutral condition, irrespective of the misophonia patient and healthy control group.

	MNI coo	rdinates		Cluster size	Z
	х	У	z	Cluster size	
L visual cortex	-48	-70	4	10895	>8
R auditory cortex	66	-12	0	3143	>8
L superior frontal gyrus	-56	4	38	231	5.3
L anterior insula	-36	28	-2	227	6.0
R amygdala	24	-22	-10	192	5.7
L anterior insula	-46	16	18	165	5.3
L inferior frontal gyrus	-30	-48	54	164	5.2



Figure 2. Brain regions with larger responses during non-neutral (misophonia and aversive) conditions than during the neutral condition, irrespective of the misophonia patient and healthy control group.



Figure 3. Significant hyperactivation in the right auditory cortex in misophonia patients compared to healthy controls, during misophonia condition compared to neutral condition.



Figure 4. Significant hyperactivation in the bilateral auditory cortex in misophonia patients, during misophonia condition compared to neutral condition.



Figure 5. Significant hyperactivation in the bilateral auditory cortex in healthy controls, during misophonia condition compared to neutral condition.

amygdala was hyperactivated in misophonia patients when the response to misophonia triggers was compared to the response to aversive stimuli. This seemed to be caused by a hypoactivation of the left amygdala in misophonia patients during the aversive condition when compared to HCs.

The BOLD response across groups was found to be higher in areas related to affective, auditory, and visual processing during non-neutral conditions when compared to the neutral condition. This finding can be explained since both the misophonia and aversive video clips are more salient, and contain more sound and movement than the neutral video clips. This comparison served as a control that the paradigm in general evoked the predicted brain responses. It was important to verify this, since the stimuli used in this paradigm had not been used in previous fMRI experiments.

Hyperactivation of the bilateral auditory cortex in misophonia patients was found when the misophonia condition was compared to the neutral condition. A possible explanation is their increased attention to auditory stimuli, which has been linked to hyperactivation of the auditory cortex (Grady et al., 1997; Jäncke et al., 1999). Further simple tests showed that both groups showed significant hyperactivation in the auditory cortex in the misophonia condition, compared to the neutral condition. This could be explained by the fact that the neutral condition was silent, as opposed to the misophonia condition.

The left amygdala was hyperactivated in misophonia patients when the misophonia condition was compared to the aversive condition. A previous experiment found that Borderline Personality Disorder (BPD) patients showed hyperactivation of the amygdala in response to emotionally aversive pictures (Herpertz et al., 2001). It was suggested that this hyperactivity reflected the intense and slowly subsiding emotions of BPD patients. Parallels with misophonia patients can be seen: they too experience more intense emotions, but in their case it is in response to trigger sounds and not to general aversive stimuli. Furthermore, hyperactivation of the amygdala has been linked to attentional functioning and vigilance (Davis & Whalen, 2001; Holland & Gallagher, 1999). Since misophonia patients are highly focused on their trigger sounds, higher vigilance during the misophonia condition is a possible explanation for the amygdalar hyperactivation.

Interestingly, we found hypoactivation of the left amygdala in misophonia patients during the aversive condition. Previous research has shown that amygdala response can be modulated by stress induction (Cousijn et al., 2010). One possible explanation for our finding is a general higher level of arousal in misophonia patients during the entire length of the paradigm, regardless of condition. If this is the case, the BOLD response in the amygdala during the fixation condition and the aversive condition, would be more similar in misophonia patients than in HCs.

Some important limitations of this experiment must be taken under close consideration. This experiment exposed the difficulties that can be encountered when trying to evoke a misophonia reaction under laboratory settings. Several misophonia patients reported



Figure 6. Significant hyperactivation in the left amygdala in misophonia patients compared to healthy controls, during misophonia condition compared to aversive condition.

that their reaction to the misophonia video clips were less forceful than a typical misophonia reaction due to it "not being real" and they could somewhat dissociate themselves from the video clips. Also, some patients experienced the repetitive sounds of the MRI scanner as misophonia triggers. This could lead to more anxiety or stress during the entire experiment, independent of the conditions. However, comparing the misophonia condition with the neutral baseline somewhat reduces this issue. Additionally, the misophonia condition consisted of four video clips with typical trigger sounds (carrot eating, grapefruit slurping, heavy breathing, and typing). Every misophonia patient has its own unique set of trigger sounds, and not all patients experienced all four video clips as triggers. Eating sounds are more common triggers than heavy breathing and typing (Schröder et al., 2013), so in order to get more misophonia reactions we could have chosen to use four eating related triggers. However, since we wanted to catch the typical reaction to a trigger, a choice was made to use a variety of sounds that matched the profile of misophonia more generally.

Furthermore, several patients were included, even if they followed one or two treatment sessions. Ideally, untreated misophonia patients who are otherwise healthy would have been used, but the number of patients registered at the AMC was too small to make such a strict selection. Finally, some patients who were invited to participate refused and reported that they felt that the experience would be too exhausting. This means that, possibly, the most severe cases of misophonia remain unexplored, and caution should be exercised in the extrapolation of the findings of current study.

#### CONCLUSION

These findings are the first direct evidence of functional abnormalities in brain responses in misophonia patients, where hyperactivation is found in the bilateral auditory cortex and in the left amygdala during a misophonia reaction. This could be explained by the hyper-focus of

	ROI	MNI coordinates				Cluster	7	n	
	NOI	х	у	Z		size	L	h	
Patients > HCs; misophonia > neutral	STC	54	-26	2	R auditory cortex	276	3.8	0.04	
Patients;	STC	-54	-34	10	L auditory cortex	2153	6.6	0.00	
misophonia > neutral		54	-26	2	R auditory cortex	2401	7.5	0.00	
HCs;	STC	-56	-8	6	L auditory cortex	1611	4.9	0.00	
misophonia > neutral		66	-12	0	R auditory cortex	1704	4.6	0.00	
Misophonia; patients > HCs	STC	-	-	-	-	-	-	n.s.	
Patients > HCs; misophonia > aversive	amygdala	-16	-2	-12	L amygdala	10	3.1	0.05	
Patients; misophonia > aversive	amygdala	-	-	-	-	-	-	n.s.	
HCs: misophonia > aversive	amygdala	-18	-4	-14	L amygdala	21	4.1	0.01	
Aversive: HCs > patients	amygdala	-16	-2	-14	L amygdala	18	3.5	0.02	
STC = superior temporal cortex. P values are small volume corrected according to corresponding ROI.									

Table 2. Brain regions that show group differences among conditions after ROI analysis.

patients on trigger sounds. Also, patients show hypoactivation of the left amygdala during a general non-misophonic aversive reaction. One possibility is that this result is caused by a general higher stress level in misophonia patients, and this should be investigated further.

Although the current study has clarified the spatial characteristics of functional abnormalities in brain responses in misophonia, this is only the first step in the neurobiological investigation of misophonia. Further research should examine if effective cognitive behavioural therapy reduces these differences in brain functioning. Long term research should focus on investigating if predictions about effectiveness of CBT can be made based on brain patterns using multivariate pattern recognition. Also, the current study highlights the role of hyper-attention to trigger sounds in misophonia. Therefore, a possible more general attentional deficit in misophonia patients should be further investigated in the future. Follow up studies should also focus on the amygdalar hypoactivation in response to non-misophonic aversive stimuli. These studies could focus on the relation between this blunted response and the degree of general distress experienced by the patient during the scanning session. In the current experiment, we also measured heart rate, and participants scored how they felt (rating on a scale their level of fear, anger, happiness, disgust, and sadness) each time after watching a video clip. The analysis of those data was beyond the scope of this report, but it should be tested to investigate a possible cause for the misophonia patients' reaction to non-misophonia aversive stimuli.

#### REFERENCES

Beesdo, K., & Lau, J. (2009). Common and distinct amygdalafunction perturbations in depressed vs anxious adolescents. Archives of ..., 66(3), 275–285. Retrieved from http://archneur. jamanetwork.com/article.aspx?articleid=482998

Boyle, D. (1996). Trainspotting.

Campion, J. (1993). The Piano.

- Cousijn, H., Rijpkema, M., Qin, S., van Marle, H. J. F., Franke, B., Hermans, E. J., ... Fernández, G. (2010). Acute stress modulates genotype effects on amygdala processing in humans. Proceedings of the National Academy of Sciences of the United States of America, 107(21), 9867–72. doi:10.1073/ pnas.1003514107
- Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. Trends in Cognitive Sciences, 3(1), 11–21. doi:10.1016/S1364-6613(98)01265-0
- Davis, M., & Whalen, P. J. (2001). The amygdala : vigilance and emotion. Molecular Psychiatry, 6, 13–34.
- Edelstein, M., Brang, D., Rouw, R., & Ramachandran, V. S. (2013). Misophonia: physiological investigations and case descriptions. Frontiers in Human Neuroscience, 7, 296. doi:10.3389/ fnhum.2013.00296
- Ferreira, G. M., Harrison, B. J., & Fontenelle, L. F. (2013). Hatred of sounds: Misophonic disorder or just an underreported psychiatric symptom? Annals of Clinical Psychiatry, 25(4), 271–274. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24199217
- Grady, C. L., Meter, J. W. Van, Maisog, J. M., Pietrini, P., Krasuski, J., & Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. Neuroreport, 8(11), 2511–2516.
- Hadjipavlou, G., & Baer, S. (2008). Selective sound intolerance and emotional distress: what every clinician should hear. Psychosomatic ..., 70, 737–740. Retrieved from http://www. psychosomaticmedicine.org/content/70/6/739.full.pdf
- Herpertz, S. C., Dietrich, T. M., Wenning, B., Krings, T., Erberich, S. G., Willmes, K., ... Sass, H. (2001). Evidence of abnormal amygdala functioning in borderline personality disorder: a functional MRI study. Biological Psychiatry, 50(4), 292–298. doi:10.1016/S0006-3223(01)01075-7

#### HYPERACTIVITY IN AMYGDALA AND AUDITORY CORTEX IN MISOPHONIA

- Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. Trends in Cognitive Sciences, 3(2), 65–73. doi:10.1016/S1364-6613(98)01271-6
- Jäncke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex : a functional magnetic resonance imaging study in human subjects. Neuroscience Letters, 266, 125–128.
- Jastreboff, P. J., & Jastreboff, M. M. (2006). Tinnitus retraining therapy: a different view on tinnitus. ORL; Journal for Oto-Rhino-Laryngology and Its Related Specialties, 68(1), 23–9; discussion 29–30. doi:10.1159/000090487
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). Principles of neural science (4th ed.). McGraw-Hill.
- Kaye, T. (1998). American History X.
- Kluckow, H., Telfer, J., & Abraham, S. (2014). Should we screen for misophonia in patients with eating disorders? A report of three cases. The International Journal of Eating Disorders. doi:10.1002/eat.22245
- Neal, M., & Cavanna, A. E. (2013). Selective sound sensitivity syndrome (misophonia) in a patient with Tourette syndrome. The Journal of Neuropsychiatry and Clinical Neurosciences, 25(1), E01. doi:10.1176/appi.neuropsych.11100235
- Nitschke, J. B., Sarinopoulos, I., Mackiewicz, K. L., Schaefer, H. S., & Davidson, R. J. (2006). Functional neuroanatomy of aversion and its anticipation. NeuroImage, 29(1), 106–116. doi:10.1016/j. neuroimage.2005.06.068
- Noë, G. (2002). Irréversible.
- Schröder, A., van Diepen, R., Mazaheri, A., Petropoulos-Petalas, D., Soto de Amesti, V., Vulink, N., & Denys, D. (2014). Diminished n1 auditory evoked potentials to oddball stimuli in misophonia patients. Frontiers in Behavioral Neuroscience, 8, 123. doi:10.3389/fnbeh.2014.00123
- Schröder, A., Vulink, N., & Denys, D. (2013). Misophonia : Diagnostic Criteria for a New Psychiatric Disorder. PloS One, 8(1). doi:10.1371/Citation
- Sztuka, A., Pospiech, L., Gawron, W., & Dudek, K. (2010). DPOAE in estimation of the function of the cochlea in tinnitus patients with normal hearing. Auris, Nasus, Larynx, 37(1), 55–60. doi:10.1016/j.anl.2009.05.001
- Thomas, K., & Drevets, W. (2001). Amygdala response to fearful faces in anxious and depressed children. Archives of ..., 58. Retrieved from http://archpsyc.jamanetwork.com/article. aspx?articleid=481847
- 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), 273–89. doi:10.1006/nimg.2001.0978

Webber, T. a, Johnson, P. L., & Storch, E. a. (2013). Pediatric

misophonia with comorbid obsessive-compulsive spectrum disorders. General Hospital Psychiatry, 18–19. doi:10.1016/j. genhosppsych.2013.10.018



Kim van Dijk

Supervisor: Jessie L. Frijling

**University of Amsterdam** 

#### ABSTRACT

Although several risk factors for the development of a posttraumatic stress disorder (PTSD) have been revealed, it is currently not possible to accurately differentiate between trauma-exposed individuals who will and those who will not develop PTSD. In addition, currently little is known about which brain dysfunctions are associated with pre-trauma or early post-trauma PTSD risk. Likewise, little is known about which brain processes may underlie development of specific PTSD symptoms. Therefore, we aimed to investigate predictive models for acute trauma-related symptoms based on structural connectivity in the brain. This study reports DTI tractography data from 22 participants who were recently exposed to a traumatic event and showed high levels of acute distress and/or acute PTSD symptoms. FA values of manually tracked (sub)structures were used for analysis. Tract integrity of the right uncinate fasciculus was found to predict acute re-experiencing scores, tract integrity of the tapetum combined with that of the left uncinate fasciculus were shown to predict acute avoidance scores, and acute hyperarousal scores were found to be predicted by tract integrity of the tapetum. Lastly, acute trauma-induced depression scores were predicted by tract integrity of the left superior cingulum. In conclusion, we found that the integrity of specific white matter (sub)structures predict trauma-related symptoms in individuals at increased risk for PTSD development acutely post-trauma. This study emphasizes the importance of investigating neural processes underlying specific symptom clusters when discrimination which brain abnormalities underlie a vulnerability for developing psychopathology after trauma-exposure in adulthood. Future studies that take factors such a previous stress-exposure into account are warranted.

**KEYWORDS** 

PTSD, Tractography, DTI, Trauma, Stress

#### **INTRODUCTION**

Approximately 4 out of 5 of people in the Netherlands experience a traumatic event that involves actual or threatened death or serious injury, or threat to physical integrity at some point in their life, after which 10% of these trauma-exposed individuals develops a post-traumatic stress disorder (PTSD; American Psychiatric Association, 2000; De Vries & Olff, 2009). PTSD is characterized by persistent re-experiencing of the trauma, avoidance of trauma-related stimuli, as well as hyperarousal, and is often accompanied by high levels of anxiety and depression (see Ginzburg, Ein-Dor & Solomon, 2010; O'Donnell, Creamer & Pattison, 2004). Although several

psychological (e.g. peri-traumatic dissociation), environmental (e.g. childhood abuse), and biological (e.g. autonomic hyperarousal) PTSD risk factors have been revealed (for a meta-analysis see Brewin, Andrews & Valentine, 2000; for a review see McFarlane, 2000), it is currently not possible to accurately differentiate between trauma-exposed individuals who will and those who will not develop PTSD. In addition, currently little is known about which brain dysfunctions are associated with pre-trauma or early post-trauma PTSD risk, and whether such brain measures could be used to differentiate between high risk and resilient trauma exposed individuals. Likewise, little is known about which brain

#### PREDICT PTSD: TRACTOGRAPHY

processes may underlie development of specific PTSD symptoms (e.g. re-experiencing, avoidance, etc.). Neuroimaging early after trauma may selectively assess pre-existent and early post-trauma measures and could therefore shed more light on the neurobiology underlying PTSD risk and symptom development.

Research has consistently shown differences in brain function when comparing PTSD patients with either healthy controls or trauma-exposed individuals without PTSD (for a review see Van Boven et al., 2009). At the core of the neurobiology underlying PTSD seems to lie a hyperresponsive amygdala, a hyporesponsive medial prefrontal cortex (including the anterior cingulate cortex (ACC)), and disrupted activation of the hippocampus (for reviews see Shin, Rauch & Pitman, 2006 and Admon, Milad & Hendler, 2013). Studies examining white matter abnormalities in PTSD have consistently reported interest in white matter bundles structurally connecting these brain regions associated with PTSD; the cingulum and the corpus callosum (for a review see Van Boven et al., 2009), and more recently also in the uncinate fasciculus. However, from cross-sectional studies it remains unclear whether findings reflect white matter integrity alterations associated with PTSD, or whether they represent a pre-existing vulnerability factor for PTSD.

As the most prominent white matter tract in the limbic system, the cingulum bundle connects the anterior cingulate gyrus in the neocortex with the entorhinal gyrus in the limbic lobe and is considered to be involved in regulating amygdala activity (Bush, Luu & Posner, 2000). Several studies have found deviations in cingulum tract integrity to be associated with a PTSD diagnosis (for a review see Ayling, Aghajani & Fouche, 2012); both decreases (Fani et al., 2012; Schuff et al., 2011; Kim et al., 2006; Kim et al., 2005; Wang et al., 2010; Zhang et al., 2011) and increases of cingulum integrity (Abe et al., 2006) have been found in PTSD patients compared to controls. Some (Yamasue et al., 2003) but not others (Abe et al., 2006; Zhang et al., 2011) have found correlations between white matter decrement of the ACC and PTSD symptom severity, in particular re-experiencing symptoms (Kim et al., 2005). Tract integrity of the cingulum has also been implicated in traumarelated depression (e.g. see Hamner, Lorberbaum & George, 1999). However, studies on trauma-related depression are fewer in number and the existence of this relation has been questioned (Fani et al., 2012). As the cingulum structurally connects the amygdala and the dACC, and it was recently proposed that pre-trauma increased reactivity in the dorsal ACC and amygdala predispose individuals for PTSD development (Admon, Milad & Hendler, 2013), the existence of pre-trauma aberrant cingulum integrity may be hypothesized to be associated with increased PTSD risk (Hammer, Lorberbaum & George, 1999; Ayling, Aghajani & Fouche, 2012). In line with this hypothesis, impaired fear extinction, which is a pre-trauma risk factor for PTSD development (Lommen et al., 2013), has been associated with stress-induced microstructural degradation of the cingulum bundle in PTSD patients (e.g. Schuff et al., 2011; Fani et al., 2012), suggesting a link between pretrauma vulnerability for PTSD development (i.e. impaired fear extinction) and cingulum integrity. In addition, other PTSD risk

factors have been associated with aberrant white matter integrity in the cingulum. For example, childhood maltreatment was found to be related to lower cingulum integrity in healthy participants (e.g. Choi et al., 2009). In a prospective study in adolescents exposed to maltreatment in childhood it was demonstrated that the lower this tract integrity, the higher the chances are of an individual to develop a major depressive episode later in life (Huang, Gundapuneedi & Rao, 2012). In addition, in a study in soldiers who returned from deployment without significant PTSD symptoms, combat stress has been shown to lead to a reduction of negative functional coupling between amygdala and dorsal ACC, but only in those individuals who perceived high levels of threat (Van Wingen, et al., 2011). This finding indicates that perception of combat-related threat, which is a risk factor for PTSD (Holbrook et al., 2001), alters functional connectivity measures potentially associated with cingulum integrity. These findings combined suggest that the experience of threatful events, either as a child or adult, may have enduring effects on both structural and functional connectivity between the amygdala and ACC, which may underlie the subsequent increased PTSD risk associated with later exposure to trauma.

The corpus callosum connects the two cerebral hemispheres and has repeatedly been the focus of (pediatric) PTSD studies, where abused children with PTSD show reduced volume in total as well as in subregions of the corpus callosum when compared to nontraumatized controls (for a review see Daniels et al., 2013). Similar to what has been found in children, adults with chronic and severe PTSD resulting from either childhood abuse or trauma-exposure in adulthood have also shown smaller callosal volumes compared to healthy controls (Kitayama et al., 2007; Villarreal et al., 2004, resp.). Likewise, higher levels of anxiety have been linked to corpus callosum abnormalties in PTSD patients (Villarreal et al., 2004); lower white matter integrity in subregions such as the most anterior section of the corpus callosum (i.e. genu) has been found to be related to anxiety levels in maltreated children (Jackowski et al., 2008). Previous reports on whether callosal abnormalities may resemble a predisposition for development of trauma-related symptoms after exposure to early trauma are conflicting. One study found a volume reduction in the genu in healthy adults exposed to high levels of early life stress (Seckfort et al., 2008). Such associations between callosal integrity and early life stress in PTSD-free individuals indicate that previous findings of reduced callosal integrity in PTSD patients might resemble a vulnerability factor for developing psychopathology opposed to a PTSD induced alternation in collosal integrity. Others however, found no abnormalities in callosal integrity or volume in either PTSD-free subjects or individuals with various psychiatric histories exposed to early traumatic stress (Choi et al., 2009 and Choi et al., 2012 resp.). It could be argued that a relation between white matter integrity and increased risk of development of PTSD symptoms might only show up when taking an important mediating factor into account. Peri-traumatic dissociation, one of the most potent risk factors for development of PTSD after trauma-exposure (Marmar, Weiss & Metzler, 1998; Ozer et al., 2008), has been show to negatively correlate with posterior callosal volume and childhood

#### VAN DIJK

dissociation scores (De Bellis et al., 2002). Moreover, the level of dissociation in healthy individuals is not only directly related to a dysfunction of hemispheric interaction through the corpus callosum (Spitzer et al., 2004), but childhood abuse is known to lead to an increase in levels of dissociation later in life (e.g. Irwin, 1994) which in turn has been shown to negatively correlate with corpus callosum volume (De Bellis et al., 2002). In sum, these findings propose a model where abnormalities in corpus callosum integrity resemble a pre-existent risk for development of trauma-related symptoms (i.e. PTSD symptoms and anxiety) through increased peri-traumatic dissociation during trauma-exposure later in life.

Apart from the cingulum and the corpus callosum, the uncinate fasciculus has recently received increasing attention in PTSD research. The uncinate fasciculus connects the amygdala and hippocampus with the ventromedial prefrontal cortex (vmPFC). In military personnel assessed both pre- and post-deployment, it was recently shown that a decrease in structural connectivity in the uncinate fasciculus from pre- to post-deployment was accompanied by a rise in PTSD symptoms, as well as reduction of hippocampal volume (Admon et al, 2012). Military personnel that came back from deployment without a significant increase in PTSD symptoms showed an increase in hippocampal volume and an increase in uncinate fasciculus integrity relative to pre-exposure measures. Of note, pre-deployment hippocampal volume or hippocampal structural connectivity with the vmPFC (i.e. uncinate fasciculus integrity) did not predict the development of PTSD symptoms during or after deployment. Therefore, the authors speculated that disrupted hippocampus-vmPFC structural connectivity is the consequence of maladaptive stress response that becomes evident only after the development of PTSD symptoms, and not when trauma exposed individuals do not develop PTSD. Likewise, it has been proposed that reduced structural and functional connectivity between the hippocampus and vmPFC is associated with a maladaptive stress response when exposed to trauma due to failed fear inhibition (Admon, Milad & Hendler, 2013), making neural abnormalities in the uncinate fasciculus an acquired feature of PTSD symptoms.

To summarize, it remains unclear how acute trauma-related symptoms relate to white matter integrity in the brain early posttrauma. If a relationship between pre-existing or early post-trauma white matter integrity and acute trauma-related symptoms exists, it is possible that this relationship is mediated by other factors such as peri-traumatic dissociation. The aim of the current study was to investigate predictive models based on structural connectivity measures of subregions of the cingulum, corpus callosum, and uncinate fasciculus for specific acute trauma-related symptoms (i.e. re-experiencing, avoidance, hyperarousal, depression, and anxiety) in recently traumatized individuals who showed high levels of acute distress and/or acute PTSD symptoms. Furthermore, we investigated whether peri-traumatic dissociation mediated any relationship between white matter integrity and acute traumarelated symptoms. Since all participants were scanned within 11 days post-trauma, we considered the DTI measures in this study to closely represent pre-existing white matter integrity (opposed to

representing white matter integrity influenced by prolonged stress exposure as is the case in PTSD patients) since structure altering processes are not yet detectable at this point in time (e.g. see Bonne et al., 2001; Wignal et al., 2004). We hypothesized that pre-existent white matter integrity in subregions of the cingulum is associated with both acute PTSD as well as depression symptoms. For the corpus callosum we hypothesized to find a predictive role of white mater integrity of the anterior projections of the corpus callosum (i.e. anterior forceps; connecting the left and right prefrontal cortices via the genu) for acute PTSD as well as anxiety scores, possibly mediated by peri-traumatic dissociation. Based on its anatomy, we expected tapetum (part of the corpus callosum, connecting right and left sided limbic structures) integrity to also be related to acute PTSD symptoms. Contrary, we expected not to find any correlations or predictive value of callosal projections which are not connecting any frontal or limbic structures, such as the parietal projections. Lastly, we hypothesized that tract integrity in the uncinate fasciculus does not predict any acute trauma-related symptoms.

#### **METHODS**

#### Population

This cross-sectional double-blind randomized placebo-controlled study included 22 recently traumatized individuals at risk of PTSD development (14 males, mean (M) age = 34.7 years, standard deviation (SD) = 11.0); a subset of the participants of a prospective double-blind randomized controlled trial; the BONDS study (Boosting Oxytocin after trauma: Neurobiology and the Development of Stressrelated psychopathology), located at the Academic Medical Center (AMC) of the University of Amsterdam. Participants were recruited from the emergency departments of the AMC and Vrije Universiteit Medical Center in Amsterdam, the Netherlands, after experiencing a potentially traumatic event according to the A1 criterion of the DSM-IV-TR (American Psychiatric Association, 2000). Participants had experienced the following types of trauma; traffic accidents (n=18), assault (n=2), and work related accidents (n=2). Only individuals between 18-65 years of age, who scored above the cutoff on screening questionnaires indicating increased risk of PTSD development, speak Dutch or English fluently and did not meet any of the exclusion criteria were eligible to participate. Participants were excluded in case of any severe or chronic systemic disease, current psychopathology (PTSD, major depression, substance abuse, psychotic disorder, bipolar disorder, or severe personality disorder), history of neurological disorder, reports of ongoing traumatization, use of certain medication (i.e. prostaglandins, ergot alkaloids,  $\beta$ -adrenergic receptor-blocking agents, and systemic glucocorticoids), impaired consciousness at time of inclusion (objectified by Glasgow coma Scale lower than 13), pregnancy or breastfeeding, or presence of metals, pacemaker or claustrophobia.

The BONDS study was approved by the Institutional Review Board of the AMC and written informed consent was given by all participants after written and oral description of the study.

#### Recruitment procedure

Potentially trauma-exposed individuals were informed of the BONDS study at the emergency departments via posters and / or verbal information provided by medical staff. Trauma patients were identified within 24 hours post-trauma through inspection of medical records for potential traumatic events that met the A1 criterion of the DSM-IV-TR. Potential participants were contacted preferably within 24-72 hours (in case of contacting difficulties up to one week) post-trauma, at which point the Trauma Screening Questionnaire (TSQ; Brewin et al., 2002) and the Peritraumatic Distress Inventory (PDI; Brunet et al., 2001), were administered after verbal consent was given (see Screening instruments). Only participants with an increased risk for developing PTSD as shown by high levels of acute distress (i.e. TSQ  $\geq$  5 and / or PDI  $\geq$  17) 1-7 days post-trauma were included in this study. The potential participant was then provided with verbal and written information on the study and invited to participate. All participants were compensated financially for their participation.

#### Experimental procedure

Within seven days post-trauma a pre-intervention assessment took place. At this time verbal and written informed consent were obtained and psychopathology-related exclusion criteria were checked by use of the MINI-PLUS (Sheehan et al., 1998; van Vliet et al., 2000). In addition, the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1995) was administered in order to evaluate acute PTSD symptoms (see Clinical interviews). Lastly, a standard MRI checklist was used to check whether a participant was eligible for participating in the imaging study. In the days between this measurement and the next appointment (i.e. 2-11 days posttrauma), participants were asked to fill out several questionnaires online, among which were the Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1983; Spinhoven et al., 1997) and the Peri-Traumatic Dissociation Questionnaire (PDEQ; Marmar, Weiss & Metzler, 1997). At the next appointment (max 11 days posttrauma), participants were placed in the scanner and several scans were made for the BONDS study, starting with the T1-weighted anatomical scan an ending with the DTI scan.

For the BONDS study, all participants received an intranasal treatment with either a placebo or oxytocin (5 puffs of 4 International Units per nostril) 30-45 minutes pre-scan. Although oxytocin is known to affect functional connectivity within and between brain regions (e.g. see Bethlehem et al., 2012), there is no indication that single dose oxytocin administration affects structural connectivity. Therefore, all analyses in this substudy are done in disregard of intranasal treatment.

#### Measures

Eligibility of trauma-patients was assessed using the TSQ (Brewin et al., 2002) and PDI (Brunet et al., 2001). The 10-item TSQ is a screening instrument that assesses acute PTSD symptoms and is known to be predictive of PTSD development (e.g. Walters, Bisson & Shepherd, 2007; Dekkers, Oldd & Näring, 2010). The PDI uses 13 items to quantify per-traumatic distress reactions and is also known to be predictive of PTSD development when administered shortly after trauma (e.g. Nishi et al., 2010).

Psychopathology-related exclusion criteria were checked using the Dutch version (van Vliet et al., 2000) of the MINI-PLUS (Sheehan et al., 1998). This widely used structured clinical interview can classify past and current DSM-IV psychiatric disorders and is easy to administer.

Acute PTSD symptom severity was assessed by use of the CAPS (Blake et al., 1995). The CAPS is the most widely used structured clinical interview for diagnosing PTSD and assessing PTSD symptom severity according to the DSM-IV-TR. It clusters symptoms as described in the DSM-IV-TR; re-experiencing (5 symptoms, range 0-40), avoidance (7 symptoms, range 0-56), and hyperarousal (5 symptoms, range 0-40). All reported symptoms are assessed based on frequency and intensity (both range 0-4), and these scores are added for total clustered symptom scores. The Dutch translation of the CAPS has shown good validity and reliability (Hovens et al., 1994).

Acute depression and anxiety levels were assessed by use of the HADS (Zigmond and Snaith, 1983). Acute post-trauma selfreports of feelings of general anxiety and depression were measured using the two corresponding subscales of the HADS. With two sets of seven items, both ranging from 0 to 21, the HADS is a wellestablished questionnaire and has shown to have good validity and reliability for the Dutch version of this questionnaire (Spinhoven et al., 1997).

Dissociative reactions during the traumatic event were measured retrospectively using the Peritraumatic Dissociation Experience Questionnaire (PDEQ; Marmar, Weiss & Metzler, 1997). Using 10-items on a 5-point scale (range 10–50) the PDEQ has shown good validity and reliability (Sijbrandij et al., 2012).

#### Imaging

Imaging data were acquired on a 3T MRI scanner (Philips Intera, Philips Medical Systems, Best, the Netherlands) with a 16-channel SENSE head coil. DTI was performed using echo-planar imaging. The DTI data were acquired using a b-value of 1000 s/mm2. Sixty coronal slices were obtained using the following sequence parameters; echo time 92 ms, repetition time 7892 ms, slice thickness 2 mm, field of view 224 x 224 x 120 mm3, and voxel size 2 x 2 x 2 mm3. The diffusion weighting was performed along 32 directions. Total DTI scan time was 7 minutes.

All data were anonymized prior to analysis. The preprocessing of the DTI data was performed using in-house developed software, written in Matlab (The MathWorks, Natick, MA). The preprocessing was executed on the Dutch Grid (www.biggrid.nl) using a web interface to the e-Bioinfra gateway (Olabarriaga, Glatard & De Boer, 2010; Shahand et al., 2011). Head motion and deformations induced by eddy currents were corrected for by an affine registration of the Diffusion Weighted Images (DWIs) to the non-diffusion weighted image. The gradient directions were corrected by the rotation

#### VAN DIJK

component of the transformation. The DWIs were resampled isotropically. Rician noise in the DWIs was reduced by an adaptive noise filtering method (Caan et al., 2010). Diffusion tensors were estimated in a non-linear least squares sense. From the tensors, fractional anisotropy (FA) maps were computed. Following DTI preprocessing, all images were inspected visually using FMRIB Software Library (FSL; Smith et al., 2004) on a Linux platform to check if preprocessing was done correctly.

Datasets were non-rigidly registered to a population-based average template. Initially, to correct for anatomical variation outside white matter, non-diffusion weighted images were registered using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) (Ashburner, 2007). Diffusion tensor datasets were warped and tensors were reoriented accordingly. Subsequently, alignment within white matter was achieved by nonrigidly registering diffusion tensor datasets using the DTI-Toolkit (DTITK) (Zhang, H., Yushkevich, Alexander & Gee, 2006). From the warped datasets, FA measures were computed.

DTI scans were transferred to a PC with a Microsoft Windows platform and processed using DTIstudio (H. Jiang and S. Mori, Johns Hopkins University and Kennedy Krieger institute, http://godzilla. kennedykrieger.org or http://lbam.med.jhmi.edu). Fiber orientation was estimated based on the largest of three eigenvalues. DTI color maps were made. Red, green, and blue colors were assigned to right-left, anterior-posterior, and superior-inferior orientations, respectively.

#### Tracts of interest selection

The cingulum, corpus callosum, and uncinate fasciculus were selected as tracks of interest for tractography purposes, as we had a strong hypothesis on associations between acute PTSD symptoms and these tracts. The cingulum and corpus callosum were divided into subregions, i.e. superior cingulum and inferior cingulum for the cingulum and forceps anterior and parietal projections for the corpus callosum.

3D tract reconstruction of tracts of interest was realized by fiber assignment using continuous tracking (i.e. FACT method; Mori et al., 1999; Xue et al., 1999) in DTIstudio. With a FA threshold of 0.15 and an inner product threshold of 0.85, angles larger than 70° were excluded during tract reconstruction. Multiple regions of interest (ROIs) were used in the reconstruction process. Based on existing anatomical knowledge of tract trajectories, ROIs were chosen in such a way that only tracts of interest would be selected. When multiple ROIs were needed for one tract, two different types of operations were used; "OR", "AND" and "NOT". Anatomical landmarks and ROIs were based on color-coded maps, see below.

Reconstruction protocols either originated from an article by Wakana et al. (2007; see also: Wakana et al., 2004), or on a-priori knowledge of tract trajectory (Oishi et al., 2011); i.e. protocols devised for the purposes of this experiment. Average FA values for distinct tracts (separated for right and left when possible) were used for further analyses.

#### Reconstruction protocols for tracts of interest

For the reconstruction protocols of the inferior cingulum, uncinate fasciculus, and forceps minor we refer to Wakana et al. (2007; tract #2, #9, and #11 resp.). Only adjustments made to these protocols, and reconstructions protocols for tracts not listed by Wakana et al. (2007) are mentioned here.

The placement of the second ROI for the superior cingulum marks an adjustment made to the protocol of Wakana et al. (2007; tract #1), see fig. 1b. Whereas these authors propose using the genu of the corpus callosum for locating the frontal part of the anterior cingulum, it can be argued that this would create a high level of inter-subject variability in defining this structure due to variations in the angle of corpus callosum relative to the base of the brain. Also, when using the genu as an endpoint for the anterior cingulum bundle, all fibers sprouting away from the cingulum before reaching this part are completely disregarded. This would be problematic since the cingulum largely consists of short u-shaped fibers connecting adjacent gyri (see also Catani & de Schotten, 2008) and therefore they do not all reach the orbitofrontal cortex. See Narr (2009) for a similar approach.

The first and second ROI for the parietal projections of the corpus callosum were placed to include the left and right hemisphere respectively, using a coronal view located about one time the thickness of the splenium posterior to the corpus callosum (fig. 2a), thereby labelling both the parietal and the occipital projections of the corpus callosum (fig. 2b). The forceps major was removed (fig. 2c) in order to leave only the parietal projections labelled (fig. 2d and 2e).

The first and second ROI for the tapetum, connecting the temporal lobes through the splenium (e.g. see Wakana et al., 2004; Sarikciogly, Ozsoy & Unver, 2007; Oishi et al., 2011), were placed using an axial plane chosen at the level of the splenium via a mid-saggital view (fig. 3a). In order to remove all non-commisural fibers, the labelled tapetum running crossing from one side of the brain to the other is also selected as a third ROI (fig. 3c). Also, fibers reaching more anterior than the tapetum were removed.

#### Statistical analysis

Descriptive statistics were computed for participant sex, age, and acute trauma-related symptoms (i.e. clustered CAPS and HADS scores). Also, distributions of all tractography results and symptoms were checked and appropriate measures were taken if necessary.

Stepwise linear regressions were used to examine correlations between and the predictive value of FA values of tracts of interest and acute trauma-related symptoms. In a second step, effects of age and sex were checked using linear regressions (enter) with FA values as the independent and CAPS/HADS scores as the dependent variable. Additionally, a bivariate correlation analysis was performed to examine the relation between peri-traumatic dissociation (i.e. PDEQ scores) and acute PTSD symptoms (i.e. clustered CAPS scores). In order to test the mediating effect of peritraumatic dissociation (i.e. PDEQ scores) on the predictive value of FA values of tracts of interest on acute trauma-related symptoms (i.e. clustered CAPS and HADS scores) a mediation analysis (independent variable = FA values, dependent variable = CAPS/ HADS scores) was performed using the PROCESS macro (Hayes & Preacher, in press) for SPSS. Lastly, in order to examine a possible acute trauma effect of on FA values, a bivariate correlation analysis was done to assess the relation between time since trauma (i.e. number of days between trauma and scan) and FA values of tracts of interest.

#### RESULTS

Descriptive statistics for participant sex, age, type of trauma, and acute trauma-related symptoms (i.e. clustered CAPS and HADS scores) as well as PDEQ scores are shown in table 1.

#### Stepwise regression analysis of FA values and CAPS/ HADS subscale scores

All assumptions (linearity of the relationship between dependent and independent variables, independence of the errors, homoscedasticity of the errors, normality of the error distribution) were met for all regressions, except for the normality of distribution for the HADS anxiety scale which did not change after transformation of the data.

CAPS re-experiencing scores correlate positively with FA values in the right uncinate fasciculus (r = 0.51, p < 0.05), the left uncinate fasciculus (r = 0.49, p = 0.01), and the forceps anterior (r = 0.48, p < 0.05). CAPS re-experiencing scores were best predicted by a model that includes the right uncinate fasciculus ( $\beta$  = 0.51, t(20) = 2.65, p < 0.05); the overall model fit was R2 = 0.26, F(1, 20) = 7.04, p < 0.05.

Secondly, CAPS avoidance scores correlated positively with FA values in the left uncinate fasciculus (r = 0.48, p < 0.05). CAPS avoidance scores were best predicted by a model that included the same structure ( $\beta$  = 0.69, t(20) = 3.64, p < 0.01) combined with the tapetum ( $\beta$  = -0.49, t(20) = -2.60, p < 0.05). The overall model fit was R2 = 0.44, F(2, 19) = 7.32, p < 0.01.

Thirdly, CAPS hyperarousal scores correlated positively with FA values in the forceps anterior (r = 0.42, p < 0.05) and the tapetum (r = 0.43, p < 0.05). CAPS hyperarousal scores were best predicted by a model that included only the tapetum ( $\beta$  = -0.43, t(20) = -2.13, p < 0.05). The overall model fit was R2 = 0.19, F(1, 20) = 4.53, p < 0.05.

Fourthly, HADS depression scores correlated positively with FA values in the left superior cingulum (r = 0.51, p < 0.01). HADS depression scores were best predicted by a model that included the same structure ( $\beta$  = -0.51, t(20) = -2.68, p < 0.05). The overall model fit was R2 = 0.26, F(1, 20) = 7.18, p < 0.05. HADS anxiety scores did not correlate with any of the FA values in the chosen tracts, no model was found to predict HADS anxiety scores.

When age and gender were also taken up into the regression analysis (enter), they showed to be no significant addition to any of the above-mentioned models. See table 2 for an overview of the regression results.

### Mediating effect of PDEQ scores on CAPS subscale scores

CAPS scores (re-experiencing, avoidance, and hyperarousal) were found to correlate with PDEQ scores (r = 0.56, p < 0.01; r = 0.59, p < 0.01; r = 0.47, p < 0.05 resp.). Also, PDEQ scores correlated positively with FA values in the right and left uncinate fasciculus (r = 0.59, p < 0.01 and r = 0.60, p < 0.01 resp.) as well as in the forceps anterior (r = 0.43, p < 0.05). However, the mediation analysis showed that there was no effect of PDEQ scores on the relation between FA values and CAPS or HADS scores.

#### FA values and time since trauma

A positive correlation between FA values in the left uncinate fasciculus and time since trauma was found (r = 0.44, p < 0.05).

#### DISCUSSION

In this study we found that the integrity of specific white matter tracts previously found to be implicated in PTSD and/or the development of PTSD were associated with acute trauma-related symptoms in individuals at increased risk for PTSD development after a recent trauma. Specifically, tract integrity of the right uncinate fasciculus was found to predict acute re-experiencing scores acutely posttrauma. Tract integrity of the tapetum and tract integrity of the left uncinate fasciculus were shown to predict acute avoidance scores. Acute hyperarousal scores were found to be predicted by tract integrity of the tapetum. Trauma-induced depression scores were predicted by tract integrity of the left superior cingulum. Additionally, tract integrity of the left uncinate fasciculus and the anterior forceps were found to correlate with (but not predict) acute re-experiencing scores. A correlation between tract integrity of the anterior forceps of the corpus callosum and hyperarousal scores was also found. Although acute PTSD symptoms were highly correlated with peritraumatic dissociation scores and PDEQ scores were related to FA values, no mediating effect on the relation with tract integrity was found, i.e. it is not likely that peri-traumatic dissociation is related to the process by which underlying tract integrity predicts acute trauma-related symptoms.

In accordance with previous research (Huang, Gundapuneedi & Rao, 2012), we found that integrity of the cingulum (albeit superior instead of inferior) predicted acute depression levels post-trauma. Contrary to our expectations we found that individuals with higher tract integrity in the left superior cingulum had higher (and not lower) acute depression scores. Others who found an increase in cingulum integrity to be related to chronic PTSD speculated that this increase (instead of decrease) is caused by the hyperresponsivity of the amygdala and represents a (temporary) secondary cortical functional activation in an effort to overcome the failure to inhibit the amygdala (Abe et al., 2006). Unlike other studies (e.g. Kim et al., 2005), we found no relation between cingulum integrity and specific PTSD symptoms. The absence of a correlation between cingulum tract integrity acutely post-trauma and acute traumarelated symptoms supports the theory of cingulum dysfunction to be induced by prolonged stress-exposure and consequently only



Figure 1. Locations of the ROIs for the superior cingulum. The first and second ROI are shown on two coronal slices (a and b resp.), with their respective locations shown on a mid-saggital slice. Also, a 3D image of the resulting labelled structure is shown (d).

shows up in chronic PTSD research or in studies were participants with and without child abuse are compared. Also, the recent theory of cingulate function underlying pre-existent psychological hyperarousal supposedly predisposing individuals for PTSD development (see Admon, Milad & Hendler, 2013) may not be directly extrapolated to structural connectivity since no relation between cingulum integrity values and hyperarousal scores was found. However, it is theoretically possible that there is a relation between pre-trauma hyperarousal scores (but not trauma-induced hyperarousal) and integrity of the cingulum bundle.

Higher tract integrity in the anterior forceps of the corpus callosum was found to be related to (but not predict) higher acute re-experiencing and hyperarousal scores. The absence of any predictive value was probably related to low correlation levels with both symptom clusters and a small sample size. Correlations between pre-existent anterior forceps integrity and re-experiencing as well as hyperarousal scores add to the mixed findings in callosal trauma-related studies. Although abnormalities in these regions are often found in (pediatric) PTSD studies, many have failed in finding any direct link with symptom severity scores (e.g. De Bellis et al., 2002; Jackowski et al., 2008). Unexpectedly, we did not find a mediating effect of peri-traumatic dissociation scores on acute trauma-related symptoms. As expected, tract integrity in the parietal projections of the corpus callosum was not found to be related to any acute trauma-related symptom scores. This finding highlights the importance of separating functionally defined substructures in DTI studies. White matter integrity in the tapetum was found to be a predictor of acute avoidance scores. Additionally, higher integrity of the tapetum was related to higher acute hyperarousal scores, and showed to be the only predictor of this PTSD subscale among these (sub)structures.

Even though higher tract integrity of the right and left uncinate fasciculus were related to higher acute re-experiencing scores, only tract integrity of the right uncinate fasciculus showed to be a good predictor of these scores. Considering its anatomy (see Von Der Heide et al., 2013) it could be expected that this tract is related to memory processes. However, finding any relation between this tract and re-experiencing scores acutely post-trauma is surprising since abnormal uncinate fasciculus integrity has been hypothesized to be stress-induced and not pre-existent (Admon et al., 2012). Contrary to our expectations we found the left uncinate fasciculus to be a good predictor for acute avoidance scores. To our knowledge, this is the first study reinforcing the theory of Von Der Heide et al. (2013) that the uncinate fasciculus supports transfer of salienceladen stimulus associations to the OFC and thus plays a vital role in guiding avoidant behavior. Also, the uncinate fasciculus was found to predict avoidance scores together with tapetum integrity.

Several issues that warrant caution in interpreting these results should be discussed. First, although brain abnormalities found this early post-trauma were considered to represent pre-existing brain characteristics (opposed to a prolonged-stress exposure state) in the current study, there are reasons to question this assumption. One is that, even though time since trauma is short, there could still be an effect of trauma-related stress exposure (in combination with the acute PTSD symptoms) on white matter integrity. However, we think this effect is small since research shows that stress-induced processes have not set in yet at this stage (e.g. see Bonne et al., 2001; Wignal et al., 2004). Therefore, the white matter integrity measured in this study is thought to represent pre-trauma measures more closely than PTSD state characteristics. Studies of TBI however, have demonstrated reduced white matter integrity in areas of axonal shearing injury acutely post (minor) head trauma. These



Figure 2. Locations of the ROIs for the parietal projections of the corpus callosum. The first and second ROI are shown on two coronal slices (a and c resp.), with the resulting labelled tracts from both steps shown on a 3D image (b and e). Also, a 3D image of the end result from a bottom up perspective is shown (d).



Figure 3. Locations of the ROIs for the tapetum. The first and second ROI are shown on an axial and a mid-saggital slices (a and c resp.). Also, 3D images of the resulting labelled structure are shown as seen from the side (d) and from the front of the brain (d).
#### VAN DIJK

effects are often not visible on conventional scans (Wilde et al., 2008), but are detectable in DTI within 24 hours post-injury (Arfanakis et al., 2002). Although we had excluded individuals with moderate to severe neurological damage (as made visible on conventional brain scans or objectified by low GCS scores), it is likely that the prevalence of mild TBI was high in our sample since a majority of the participants had experienced deceleration injury (N=18 for traffic accidents), a common cause of TBI. To check our assumption of measuring pre-existent white matter integrity, we examined the relation between DTI measures and time since trauma, assuming that white matter integrity changes caused by mild TBI would be more apparent closer to the time of trauma. It was found that tract integrity only in the left uncinate fasciculus was positively related to time since trauma, implying that the FA-values of left the uncinate fasciculus in our samples may not be an accurate representation of pre-trauma white matter integrity. This is not surprising considering the location of this tract and its proven implication in TBI cases (e.g. Niogi et al., 2010). Since no other structure was shown to be related to time since trauma, we concluded that these findings could, for the purpose of the current study, be considered as a preexistent measure of white matter integrity. Second, all participants in this study had shown increased risk for PTSD development and therefore belong to a strongly biased group. It is possible that some of the expected results (i.e. predictive role of cingulum integrity for acute PTSD symptoms) can only be found when looking at a more general population with variation in the level of PTSD risk. Lastly, based on previous studies, it can be hypothesized that the white matter integrity of our studied tracts (e.g corpus callosum) is influenced by having experienced childhood trauma. Further studies are needed to unravel the exact mechanism though which childhood trauma influences white matter integrity and whether the relationship between childhood trauma and acute PTSD symptoms may be mediated by this integrity. Similarly, a history of PTSD or other psychopathology may also have influenced white matter integrity in the studied tracts (Thomason & Thompson, 2011). Since all individuals in this study belonged to a strongly biased group

Table 1. Demographic and clinical characteristics.

Variable	Mean (SD)
Participants (N=22, 14 males)	
Age	34.7 (11.0)
Traffic accidents (N=18)	
Assault (N=2)	
Work-related accidents (N=2)	
CAPS	
Re-experiencing	15.5 (9.1)
Avoidance	8.8 (7.9)
Hyperarousal	15.2 (8.2)
HADS	
Depression	8.0 (2.2)
Anxiety	12.5 (2.4)
PDEQ	25.9 (8.9)

(i.e. high risk for PTSD) it is likely that the prevalence of childhood abuse and previously experienced PTSD is higher in our studied group than in other populations. Both of these factors are known to influence structural connectivity in the brain (e.g. see Hart & Rubia, 2012). We therefore cannot accurately conclude whether our results reflect a 'clean' pre-existent measure of PTSD vulnerability, the influence of early life stress, the effects of a previous PTSD, or a combination of these factors.

Table 2. Results of the linear regressions for predicting CAPS/HADS scores based on FA values.

	CAPS			HADS
	Re-experiencing	Avoidance	Hyperarousal	Depression
Correlating				left superior cingulum**
FA values				
	right / left uncinate	left uncinate		
	fasciculus**	fasciculus*		
	forceps anterior*		forceps anterior*	
			tapetum*	
Prediction model				left superior cingulum*
	right uncinate fasciculus*	Left uncinate		
		fasciculus**		
		tapetum*	tapetum*	
p < 0.05; **p < 0.01	1			

#### CONCLUSION

This study shows that the structural integrity of several white matter (sub)structures predict specific trauma-related symptoms early posttrauma. These findings highlight the importance of tractography in neuroimaging, since certain areas such as the tapetum had not been studied due to technical challenges but are found to predict 2 out of 3 PTSD symptom clusters acutely post-trauma in individuals at risk for PTSD. Also, this study emphasizes the importance of investigating neural processes underlying specific symptom clusters when discrimination which brain abnormalities underlie a vulnerability for developing psychopathology after trauma-exposure in adulthood. In conclusion, neuroimaging studies early after trauma can selectively assess pre-existent and early post-trauma measures, and shed more light on the neurobiology underlying PTSD risk and symptom development. However, future studies that take factors such a previous stress-exposure (e.g. (childhood) trauma, history of psychopathology) into account are warranted, to further investigate the precise relationship between white matter integrity, (traumatic) stress -exposure and PTSD risk. Preferably, this study should also be repeated in a low risk trauma-exposed population.

#### REFERENCES

- Abe, O. et al. (2006). Voxel-based diffusion tensor analysis reveals aberrant anterior cingulum integrity in posttraumatic stress disorder due to terrorism, Neuroimaging, 146, 231-242.
- Admon, R. et al. (2012). Stress-induced reduction in hippocampal volume and connectivity with the ventromedial prefrontal cortex are related to maladaptive responses to stressful military service, Human Brain Mapping, 34, 11, 2808-2816.
- Admon, R., Milad, M. R., & Hendler, T. (2013). A causal model of posttraumatic stress disorder: disentangling predisposed from acquired neural abnormalities, Trends in Cognitive Sciences, 17, 7, 337-347.
- American Psychiatric Association, 2000. Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.
- Arfanakis, K. et al. (2002). Diffusion tensor MR imaging in diffuse axonal injury, American Journal of Neuroradiology, 23, 5, 794-802.
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm, NeuroImage, 38, 1, 95-113.
- Ayling, E., Aghajani, M., & Fouche, J.-P. (2012). Diffusion tensor imaging in anxiety disorders, Current Psychiatry Reports, 14, 197-202.
- Bethlehem, R. A., Van Honk, J., Auyeung, B., & Baron-Cohen, S. (2012). Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies, Psychoneuroendocrinology, 38, 7, 962-974.
- Blake, D. D. et al. (1995). The development of a Clinician-Administered PTSD Scale, Journal of Trauma Stress, 8, 1, 75-

90.

- Bonne, O. et al. (2001). Longitudinal MRI study of hippocampal volume in trauma survivors with PTSD, American Journal of Psychiatry, 158, 8, 1248-1251.
- Brewin, C. R. et al. (2002). Brief Screening instrument for posttraumatic stress disorder, British Journal of Psychiatry, 181, 158-162.
- Brewin, C. R., Andrews, B., & Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in traumaexposed adults, Journal of Consulting and Clinical Psychology, 68, 5, 748-766.
- Brunet, A., Weiss, D.S., Metzler, T.J., et al. (2001). The Peritraumatic Distress Inventory: A proposed measure of PTSD criterion A2. American Journal of Psychiatry, 158, 1480-1485.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex, Trends in Cognitive Sciences, 4, 6, 215-222.
- Caan, M. W. A., et al. (2010). Adaptive noise filtering for accurate and precise diffusion estimation in fiber crossings. The Medical Image Computing and Computer Assisted Intervention Society, 13, 167-174.
- Catani, M., & de Schotten, M. T. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections, Cortex, 44 (8), 1105-1132.
- Choi, J., Jeong, B., Polcari, A., Rohan, M.L. & Teicher, M.H. (2012). Reduced fractional anisotropy in the visual limbic pathway of young adults witnessing domestic violence in childhood, Neuroimage, 59, 2, 1071-1079.
- Choi, J., Jeong, B., Rohan, M. L., Polcari, A. M., & Teicher, M. H. (2009). Preliminary evidence for white matter tract abnormalities in young adults exposed to parental verbal abuse, Biological Psychiatry, 65, 3, 227-234.
- Daniels, J. K., Lamke, J-P., Gaebler, M., Walter, H., & Scheel, M. (2013). White matter integrity and its relationship to PTSD and childhood trauma – systematic review an meta-analysis, Depression and Anxiety, 30, 3, 207-216.
- De Bellis, M. D. et al. (2002). Brain structures in pediatric maltreatment-related posttraumatic stress disorder: A sociodemographically matched study, Society of Biological Psychiatry,52, 1066-1078.
- De Vries, G. J., & Olff, M. (2009). The lifetime prevalence of traumatic events and posttraumatic stress disorder in the Netherlands. Journal of Traumatic Stress, 22, 259-267.
- Dekkers, A. M. M., Oldd, M., & Näring, G. W. B. (2010). Identifying persons at risk for PTSD after trauma with TSQ in the Netherlands. Community Mental Health Journal, 46, 1, 20-25.
- Fani, N. et al. (2012). Attention bias toward threat is associated with exaggerated fear expression and impaired extinction in PTSD. Psychological Medicine, 42, 3, 533 543.

- Ginzburg, K., Ein-Dor, T, & Solomon, Z. (2010). Comorbidity of posttraumatic stress disorder, anxiety and depression: A 20year longitudinal study of war veterans, Journal of Affective Disorders, 123, 1-3, 249-257.
- Hammer, B. H., Lorberbaum, J. P., & George, M. S. (1999). Potential role of the anterior cingulate cortex in PTSD: Review and hypothesis. Depression and Anxiety, 9, 1-14.
- Hart, H., & Rubia, K. (2012). Neuroimaging of child abuse: a critical review, Frontiers in Human Neuroscience, 6, 52.
- Hovens, J. E. et al. (1994). The assessment of posttraumatic stress disorder: with the Clinician Administered PTSD Scale: Dutch results. Journal of Clinical Psychology, 50, 3, 325-340.
- Irwin, H. J. (1994). Proneness to dissociation and traumatic childhood events, The Journal of Nervous and Mental Disease, 182, 8, 456-460.
- Holbrook, T.L., Hoyt, D.B., Stein, M.B., & Sieber, W.J. (2001). Perceived threat to life predicts posttraumatic stress disorder after major trauma: Risk factors and functional outcome, The Journal of Trauma: Injury, Infection, and Critical Care, 51, 2, 287-293.
- Huang, H. Gundapuneedi, T., & Rao, U. (2012). White matter disruptions in adolescents exposed to childhood maltreatment and vulnerability to psychopathology, Neuropsychopharmacology, 37, 12, 2693-2701.
- Jackowski, A. P. et al. (2008). Corpus callosum in maltreated children with posttraumatic stress disorder: A diffusion tensor imaging study, Psychiatry Research: Neuroimaging, 162, 256-261.
- Kim, M. J. et al. (2005). Disrupted white matter tract integrity of anterior cingulate in trauma survivors, NeuroReport, 16, 10, 1049-1053.
- Kim, S. J. et al (2006). Asymmetrically altered integrity of cingulum bundle in posttraumatic stress disorder, Neuropsychobiology, 54, 120-125.
- Kitayama, N., Brummer, M., Hertz, L., & Quinn, S. (2007). Morphologic alterations in the corpus callosum in a b u s e related posttraumatic stress disorder a preliminary study, Journal of Nervous and Mental disease, 195, 1027-1029.
- Lommen, M.J.J., engelhard, R.M., Sijbrandij, M., van den Hout, M.A. & Hermans, D. (2013). Pre-trauma individual differences in extinction learning predict posttraumatic stress, Behaviour Research and Therapy, 51, 2, 63-67.
- Marmar, C. R., Weiss, D.S., & Metzler, T. J. (1998). The peritraumatic dissociative experiences questionnaire. In Wilson JP et al (eds), Assessing psychological trauma and PTSD. NY: Guildford Press.
- McFarlane, A. C. (2000). Posttraumatic stress disorder: A model of the longitudinal course and the role of risk factors, Journal of Clinical Psychiatry, 61, 5, 15-23.

- Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C.M. (1999). Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging, Annals of Neurology, 45, 265– 269.
- Narr, K. (2009). Identifying the cingulate gyrus component of the cingulum in DTIStudio using a multiple ROI approach. Retrieved from http://www.loni.ucla.edu/~narr/protocol.php?q=dti\_ cingulum
- Niogi, S. N., Mukherjee, P., Bazarian, J. J. (2010). Diffusion tensor imaging of mild traumatic brain injury, Journal of Head Trauma Rehabilitation, 25, 4, 241-255.
- Nishi, D., et al. (2010). Peritraumatic Distress Inventory as a predictor of post-traumatic stress disorder after a s e v e r e motor vehicle accident. Psychiatry and Clinical Neuroscience, 64, 2, 149-156.
- O'Donnell, M. L., Creamer, M., & Pattison, P. (2004). Posttraumatic stress disorder and depression following trauma: Understanding comorbidity, The American Journal of Psychiatry, 161, 8, 1390-1396.
- Oishi, K., Faria, A., van Zijl, P. C. M., & Mori, S. (2011). MRI atlas of human white matter. (Elsevier B. V., eds). UK: CPI Group. First edition 2005.
- Olabarriaga, S. D., Glatard, T., & De Boer, P. T. (2010). A virtual laboratory for medical image analysis, IEEE Transactions on Information Technology in Biomedicine, 14, 4, 979-85.
- Ozer, E. J., Best, S.R., Lipsey, T.L., & Weiss, D.S. (2008). Predictors of posttraumatic stress disorder and symptoms in adults: A metaanalysis, Psychological Trauma: Theory, Research, Practice, & Policy, S, 13–36.
- Schuff, N. et al (2011). Patterns of altered cortical perfusion and diminished subncortical integrity in posttraumatic stress disorder: An MRI study, NeuroImage, 54, S62-S68.
- Seckfort, D.L. et al. (2008). Early life stress on brain structure and function across the lifespan: a preliminary study, Brain Imaging and Behavior, 2, 1, 49-58.
- Shahand, S., et al. (2011). Front-ends to biomedical data analysis on grids, Proceedings of the HealthGrid Conference, 27-28.
- Sheehan, D. V. et al. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. Journal of Clinical Psychiatry, 59, 20, 22-33.
- Shin, L. M., Rauch, S. L., & Pitman, R. K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD, Annals of the New York Academy of Sciences, 1071, 67-79.
- Sijbrandij, M., et al. (2012). The structure of peritraumatic dissociation: a cross validation in clinical and nonclinical samples, Journal of Traumatic Stress, 25, 4, 475-479

Smith, S. M. et al. (2004). Advances in functional and structural MR

image analysis and implementation as FSL, NeuroImage, 23, 208-219.

- Spinhoven, P. et al. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects, Psychological Medicine, 27, 2, 363-370.
- Teicher, M. H., Samson, J. A., Yi-Shin Sheu, M. A., Polcari, A., & McGreenery, C. E. (2010). Hurtful words: Association of exposure to peer verbal abuse with elevated psychiatric symptom scores and corpus callosum abnormalities, American Journal of Psychiatry, 167, 12, 1464-1471.
- Thomason, M. E., & Thompson, P. M. (2011). Diffusion imaging, white matter, and psychopathology, Annual Review of Clinical Psychology, 7, 63-85.
- Van Boven et al. (2009). Advances in neuroimaging of traumatic brain injury and posttraumatic stress disorder. Journal of Rehabilitation Research & Development, 46, 6, 717-756.
- Van Vliet, I. H., Leroy, H.,, & van Megen, H. (2000). MINI Plus: MINI International Neuropsychiatric Interview: Dutch version 5.0. LUMC.
- Van Wingen, G.A., Geuze, E., Vermetten, E., & Fernández, G. (2011). Perceived threat predicts the neural sequelae of combat stress, Molecular Psychiatry, 16, 664-671.
- Villarreal G. et al. (2004). Reduced area of the corpus callosum in posttraumatic stress disorder, Psychiatry Research, 131. 3, 227-235.
- Von Der Heide, R. J., Skipper, L. M., Klobusicky, E., & Olson, I. R. (2013). Dissecting the uncinate fasciculus: disorders, controversies and a hypothesis, Brain, 136, 6, 1692-1707.
- Wakana, S. et al (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter, NeuroImage, 36, 630-644.
- Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C. M., & Mori, S. (2004). Fiber tract-based atlas of human white anatomy, Radiology, 230, 77-87.
- Walters, J.T., Bisson, J.I., & Shepherd, J.P. (2007). Predicting posttraumatic stress disorder: validation of the Trauma Screening Questionnaire in victims of assault, Psychological Medicine, 37, 1, 143-150.
- Wang, H.-H. et al. (2010). Psychopathological, biological, and neuroimaging characterization of posttraumatic stress disorder in survivors of a severe coalmining disaster in China, Journal of Psychiatric Research, 44, 385-392.
- Wignal, E. L. et al. (2004). Smaller hippocampal volume in patients with recent-onset posttraumatic stress disorder. Society of Biological Psychiatry, 56, 832-836.
- Wilde, E. A. et al. (2008). Diffusion tensor imaging of acute mild traumatic brain injury in adolescents, Neurology, 70, 948-955.

Xue, R., van Zijl, P.C.M., Crain, B.J., Solaiyappan, M., Mori, S.

(1999). In vivo three-dimensional reconstruction of rat brain axonal projections by diffusion tensor imaging, Magnetic Resonance in Medicine, 42, 1123–1127.

- Yamasue, H. et al. (2003). Voxel-based analysis of MRI reveals anterior cingulate gray-matter volume reduction in posttraumatic stress disorder due to terrorism. Proceedings of the National Academy of Sciences of the United States of America, 100, 9039-9043.
- Zhang, L. et al. (2011). Different white matter abnormalities between the first-episode treatment-naïve patients with posttraumatic stress disorder and generalized anxiety disorder without comorbid conditions. Journal of Affective Disorders, 133, 294-299.
- Zhang, H., Yushkevich, P. A., Alexander, D. C., & Gee, J. C. (2006). Deformable registration of diffusion tensor MR images with explicit orientation optimization. Medical Image Analysis, 10, 5, 764–785.
- Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale, Acta Psychiatrica Scandinavica, 67, 6, 361-370.



# **Experience-dependent suppression** of mu- and beta-power in the infant motor cortex while observing others' actions

**Rianne van Rooijen** Supervisor: Janny Stapel University of Amsterdam

#### **ABSTRACT**

Action observation and action execution are closely linked, as they lead to similar cortical motor activity. This similar activity probably arises from the mirror system, which supports action understanding and interpretation. The link between action observation and execution might develop by gaining motor experience with an action. Previous research in infants indicated that more experience with an action leads to more cortical motor activation during the observation of that action. However, an alternative explanation might be that these differences are due to brain maturation accompanying action experience. This study aimed to rule out this alternative explanation by testing two groups of infants with no prior walking experience, yet one group received a one-week walking training by using a walker prior to test. The results indicated that training did not show to have a significant impact on the cortical motor response to action observation, as the groups did not differ in their cortical motor response to walking. Therefore, we cannot rule out that previously found differences might be due to brain maturation. Future research using longer training periods might help disentangle whether one's own action experience changes the processing of the observation of actions. The effect of action experience on action observation might only arise after extensive periods of training.

#### **KEYWORDS**

Action Observation, Action Experience, Mirror System

#### **INTRODUCTION**

Infants learn many things during their development, of which the acquisition of motor skills is one. In early stages of development infants merely observe other people's actions, yet they gradually become able to perform these actions themselves. Recent studies investigated whether this gradual acquisition of motor skills alters an individual's representation of the action of others. This study examined specifically whether training infants in an action, i.e. walking, alters their motor activation during the observation of this action. It is important to develop reliable representations of others' actions as it is suggested that such representations play a critical role in understanding and predicting the actions of others (Sebanz & Knoblich, 2009).

of that action.

It is hypothesized that action observation and action execution

The link between action observation and action execution

similar activation in the precentral motor cortex (Grèzes & Decety, 2001; Hari, Forss, Avikainen, Kirveskari, Salenius & Rizzolatti, 1998). It is suggested that similarities in brain activation during action observation and action execution arise from the mirror system (Rizzollati & Craighero, 2004). First evidence for a mirror system was provided by Gallese and colleagues (1996), who found that neurons in the monkey ventral premotor cortex are activated during both the execution of a specific action and the observation of another individual performing that same action. Thus, the mirror system seems to support action recognition and the interpretation of the observed action by activating one's own motor representation

are closely linked. Both action observation and execution lead to

#### EEG-POWER SUPPRESSION DURING ACTION OBSERVATION

potentially develops by gaining motor experience with a specific action. Recent studies indicated that the processing of others' actions is affected by one's own motor skills (Calvo-Merino, Glaser, Grèzes, Passingham & Haggard, 2005; Calvo-Merino, Grèzes, Glaser, Passingham & Haggard, 2006; Cross, Hamilton & Grafton, 2006; Haslinger, Erhard, Altenmüller, Schroeder, Boecker & Ceballos-Baumann, 2005). Calvo-Merino and colleagues (2005) investigated this issue by conducting an experiment with two groups of expert dancers, one group trained in professional ballet and the other group trained in capoeira, and a control group of inexperienced dancers. Participants watched videos of dance moves specific for professional ballet and capoeira while lying in an fMRI-scanner. Experts showed greater activation of the human mirror system during the observation of movements specific to their own dance style compared to movements not incorporated in their own motor repertoire. The control group did not show a difference in activation of the mirror system during the observation of both dance styles. Thus, one's own experience with a specific movement seems to alter the representation of observing someone else making that same movement.

However, an alternative explanation for these results might be that the experts have a greater visual familiarity for movements specific to their own dance style. To rule out this explanation Calvo-Merino and colleagues (2006) conducted a follow-up study with male and female professional ballet dancers. Both males and females have visual experience with dance moves of either gender, yet they are only trained in the moves specific for their own gender. The participants watched videos of male, female and gendercommon dance moves while lying in an fMRI-scanner. Participants showed higher activation in brain areas associated with the human mirror system while observing movements specific for their own gender and gender-common movements compared to movements specific for the opposite gender. Thus, it seems that the observation of an action is activating a motor representation specifically and this response is dependent on one's own experience with performing that action. Yet, these studies merely focused on expert movements, acquired by just a small proportion of people after extensive periods of training. The question arises whether the same processes underlie the natural acquisition of motor skills.

To address this question, research has to be conducted with young infants who are in the middle of developing their motor skills. Recent developmental studies showed that also in infants action observation leads to activation of cortical motor areas (e.g. Marshall, Young & Meltzoff, 2011; Meyer, Hunnius, van Elk, van Ede & Bekkering, 2011). Cortical motor activation in infants is usually assessed by means of frequency analysis of the electrophysiological signal. Just as in adults (Caetano, Jousmaki & Hari, 2007; Muthukumaraswamy & Johnson, 2004), motor activation is indicated by a power decrease in the mu- and beta-frequency band over cortical motor sites (Southgate, Johnson, Osborne & Csibra, 2009; Meyer et al., 2011).

The link between the natural acquisition of motor skills and

motor activation during action observation in infants was directly investigated by van Elk and colleagues (2008). In their study, 14to 16-month old infants saw short movies of walking and crawling infants while the infants' EEG was recorded simultaneously. Stronger mu- and beta-suppression was found for the observation of crawling compared to walking videos. This indicates a stronger motor activation during observation of the crawling videos, which is the movement the infants had most experience with. The link between mu-power during action observation and one's own experience was also indicated by a significant correlation between the difference in mu-power during the observation of both movements and the infant's own crawling experience. Infants with more crawling experience showed a greater suppression of mu-power for the observation of crawling compared to walking videos. A similar correlation was found between beta-power during action observation and crawling experience.

However, an alternative explanation for these results might be that infants with more crawling experience are further in their brain maturation and therefore show different activation patterns. The current study aimed to rule out this alternative explanation by testing two groups of infants who did not have any prior walking experience, yet one group received a walking training prior to testing. Hence, we tested 11-month-old infants who did not have prior walking experience, yet they already had some crawling experience. Half of the participants were trained in walking by using a walker for one week, such that they would gather walking experience. The other half of the participants did not receive such a training.

A training duration of one week was chosen because previous research has shown that one week of training can alter activation in cortical motor areas. That is, Paulus and colleagues (2012) showed that one week of training can affect activation in the mufrequency band over cortical motor sites. In this study, 8-monthold infants received a training in shaking a rattle which produced a specific sound. They also heard another sound during this week, but this sound was not related to any action the infants performed themselves. Results indicated that at test the infants showed greater mu-suppression while listening to the sound of the rattle compared to the other sound heard during training and a complete novel sound.

An important difference between the study conducted by Paulus et al. (2012) and the current study is that the infants who received the rattle training were already able to shake a rattle, thus the action was already part of their motor repertoire prior to training. The training served to build action-effect associations. In contrast, infants who received the walking training were not yet able to walk, thus training added a new action to their motor repertoire. Hence, this study investigated whether one week of training results in similar effects on motor activation for actions which were and were not part of the infants' motor repertoire prior to training.

Based on the studies by van Elk et al. (2008) and Paulus et al. (2012) we hypothesized that the infants in the control group would show a reduction in mu- and beta-power for the observation of crawling compared to walking movements, as they only have

#### VAN ROOIJEN

experience with crawling. The infants in the training group would show a reduced difference in mu- and beta-power during the observation of both movements, as they have experience with both walking and crawling.

Furthermore, we expected to find a correlation between the difference in mu- and beta-power for observing crawling compared to walking movements and the amount of motor experience, similar to the studies of Calvo-Merino et al. (2005) and van Elk et al. (2008). The infants with more crawling experience were expected to show a greater difference in mu- and beta-power during the observation of crawling compared to walking movements. Moreover, the infants in the training group who received most training would show the largest reduction in the difference in mu- and beta-power while observing crawling compared to walking movements.

#### **METHODS**

#### Participants

Thirteen 11-month-old full-term infants participated in this experiment; six in the training group (mean age = 11 months and 1 day, SD = 10 days; 1 girl) and seven in the control group (mean age = 11 months and 9 days, SD = 10 days; 3 girls). An additional 15 infants were tested but excluded from analysis due to EEG recordings of insufficient quality resulting from the inability to lower impedances because of the infants' limited patience (n=8) or a lack of at least 10 movement- and artifact-free trials per condition (n=7).

The infants were recruited from the database of the Baby Research Center Nijmegen, which consists of families willing to participate in child studies. Infants were selected to take part in this experiment if they met several requirements. First, the infants should be capable of crawling by using hands and knees while lifting their belly. This was to ensure that the infants themselves had experience with the movement shown in the crawling videos. Second, the infants should not yet be able to walk independently. Last, they were not allowed to have a walker at home. These last two criteria were to ensure that the infants did not have any walking experience prior to the study.

Parents signed an informed consent form prior to the study and the infants received a gift or a monetary reward in appreciation for their participation. The study was approved by the local ethics committee Arnhem Nijmegen.

#### Stimuli

For the present study we used the exact same stimulus videos as used by van Elk et al. (2008). These videos displayed a walking or crawling infant, recorded at the Baby Research Center Nijmegen. Parents were informed about the purpose of this study and gave their consent for the use of these videos in the present experiment. In the videos, an infant walks or crawls for approximately 2 m from the left to the right side of the visual scene. Figure 1 displays an example frame from a crawling and a walking video. The videos were flipped horizontally such that both leftward and rightward movements could be shown. This resulted in a total stimulus set of 48 videos, which consisted of 15 walking and 9 crawling videos presented in both a leftward and rightward direction. The videos ranged from 1840 ms to 6040 ms in duration.



Figure 1. Crawling and walking stimuli. On the left a frame from one of the crawling videos and on the right a frame from one of the walking videos.

#### Procedure

The infants were randomly assigned to either the training or the control group. One week before the EEG session, the infants in the training group received a walker (Chicco Band, Artsana group, Grandate) at home. The walker is displayed in Figure 2. The parents were instructed to let the infant practice walking in the walker for 10 minutes a day, for seven consecutive days. Parents filled in a training scheme to indicate when and how long the infant had actually trained with the walker. The infants in the control group did not receive this training with a walker but only participated in the EEG-experiment. One week before the test day, all parents received a questionnaire to indicate at which date their infant started crawling, if possible based on diaries or agendas. This information was used to calculate how many days of crawling experience the infants had at the moment of testing.



Figure 2. The walker used in this study.

During the actual EEG-experiment the infants were seated in a car seat in front of a computer screen at a distance of approximately 60 cm. Blocks of walking and crawling videos were randomly presented to the infants, each block lasting between 7.5 and 14 seconds. During a block of videos, leftward and rightward movements were shown in alternation, each block starting with a rightward movement and ending with a leftward movement. With the start of each video, EEG markers were sent such that video onsets could be traced back in the EEG recordings. Attention getters were used to regain attention when the infant looked away from the screen. These attention getters were visually attractive videos with accompanying sounds. The experiment continued until the infant did not look at the screen anymore and we were unable to regain attention with attention getters or until the maximum of 30 blocks was attained. During the course of the experiment the infant's EEG was recorded and the behaviour of the infant was video-recorded.

#### EEG recordings

Infants' EEG was recorded using an infant-sized EEG cap with 32 electrodes. The Ag/AgCl electrodes were placed in an actiCap (Brain Products, Munich). The electrodes were referenced online to electrode FCz. We used a 32-channel BrainAmp DC EEG amplifier (Brain Products, Munich) using a band-pass of 0.1 - 1000 Hz at a sampling rate of 500 Hz. We strived to keep all impedances below 60 k $\Omega$ .

#### EEG data analysis

The EEG data was analyzed using FieldTrip software, an open source Matlab toolbox developed by the FC Donders Centre for Cognitive Neuro-imaging (Oostenveld, Fries, Maris & Schoffelen, 2011).

First, gaze and movement behaviour of the infants was coded based on the video recordings. The EEG data collected during stimulus presentation was segmented into epochs of 1000 ms in which the infant looked at the screen, which formed the trials. Trials in which the infant moved were excluded from the analysis. Subsequent artifact rejection was based on visual inspection of the trials. As a result 18% of the trials were excluded from analysis. Only infants with more than 10 walking and 10 crawling trials were selected for further analysis. This left an average of 30 crawling trials and 29 walking trials per infant.

Next, a frequency analysis was performed on the remaining trials with a range from 2 to 40 Hz as frequencies of interest. For each infant, the power estimates of these frequencies during the observation of crawling and walking videos were calculated by analyzing the trials using the multitaper frequency transformation method with Hanning tapers. The resulting power values were averaged over trials for each action. Then, the mu-power per infant was calculated by averaging the power values of the frequency range of 6 to 9 Hz. A similar calculation was performed for the beta-power with a frequency range of 17 to 19 Hz.

Subsequently, a 2 (training vs. control) x 2 (crawling vs. walking) repeated measures ANOVA was used to assess the difference in mu-power while observing crawling compared to walking videos over electrode C3 and C4, the cortical motor sites. A similar analysis was performed for the beta-power.

Last, the correlation between crawling experience and the individual's difference in mu- and beta-power for crawling compared to walking videos was calculated per group to assess the relation between motor experience and effects in the mu- and beta-frequency bands. Moreover, in the training group the correlation between the amount of training and the individual's difference in mu- and betapower for crawling compared to walking videos was calculated to assess the effect of training on motor activation.

#### RESULTS

#### Mu-frequency power

A repeated measures ANOVA was used to investigate whether the groups (training vs. control) showed a different difference in the mufrequency power (6-9 Hz) for the observation of the two movements (crawling vs. walking). No main effect of movement on mu-power was found (F(1,11)=.85, ns). More importantly, there was no interaction effect between movement and group on mu-power (F(1,11)=.02, ns). Thus, there were no differences in mu-power while observing crawling compared to walking movements, irrespective of group. These results are shown in Figure 3.



Figure 3. Mu-power while observing crawling and walking movements. Displayed is the normalized mu-power as a function of observed movement (crawling; walking), separated by group. Error bars indicate the standard error of the mean.

Next, we investigated the relation between the individual experience with the observed movements and the responses in the mu-frequency range during the observation of these movements. No significant correlations were found between the amount of crawling experience and the difference in mu-power while observing crawling compared to walking movements, in neither of the groups (control group: r=.55, ns; training group: r=-.28, ns). These correlations are displayed in Figure 4. In the training group, controlling for the amount of walking training yielded a still insignificant correlation between crawling experience and the difference in mu-power while observing crawling compared to walking movements (r=.37, ns).



Figure 4. Correlations between crawling experience and difference in mu-power. The correlation between crawling experience and the normalized difference in mupower while observing crawling compared to walking movements is displayed for the control group (A) and the training group (B).

#### VAN ROOIJEN

Furthermore, no significant correlation was found between the difference in mu-power while observing crawling compared to walking movements and the amount of walking training (r=.62, ns). Controlling for crawling experience resulted in a similar, insignificant correlation between the difference in mu-power while observing crawling compared to walking movements and the amount of walking training (r=.65, ns). This correlation is displayed in Figure 5.



Figure 5. Correlation between walking experience and difference in mu-power. The correlation between walking experience (i.e. the amount of walking training) and the normalized difference in mu-power while observing crawling compared to walking movements is displayed.

#### Beta-frequency power

A similar ANOVA as for the mu-frequency range was also conducted to analyse potential effects of training on responses to action observation in the beta-frequency range (17-19 Hz). No main effect of movement on beta-power was found (F(1,11)=.10, ns). Moreover, there was no interaction effect between movement and group on beta-power (F(1,11)=.29, ns). Thus, there were no differences in beta-power while observing crawling compared to walking movements, irrespective of group. These results are shown in Figure 6.



Figure 6. Beta-power while observing crawling and walking. Displayed is the normalized beta-power as a function of observed movement (crawling vs. walking), separated by group. Error bars indicate the standard error of the mean.

Comparable to the analyses in the mu-frequency range, we investigated the relation between individual differences in experience with the observed actions and the effect on responses in the beta-frequency range during the observation of these movements. Again, no significant correlations were found between the amount of crawling experience and the difference in beta-power while observing crawling compared to walking movements, in neither of the groups (control group: r=-.14, ns; training group: r=-.61, ns). These correlations are displayed in Figure 7. In the training group, controlling for the amount of walking training did yield a similar, insignificant correlation between crawling experience and the difference in beta-power while observing crawling compared to walking movements (r=-.50, ns).



Figure 7. Correlations between crawling experience and difference in beta-power. The correlation between crawling experience and the normalized difference in beta-power while observing crawling compared to walking movements is displayed for the control group (A) and the training group (B).

Furthermore, no significant correlation was found between the difference in beta-power while observing crawling compared to walking movements and the amount of walking training (r=.42, ns). Controlling for crawling experience resulted in a still insignificant correlation between the difference in beta-power while observing crawling compared to walking movements and the amount of walking training (r=-.09, ns). This correlation is displayed in Figure 8.



Figure 8. Correlation between walking experience and difference in beta-power. The correlation between walking experience (i.e. the amount of walking training) and the normalized difference in beta-power while observing crawling compared to walking movements is displayed.

#### EEG-POWER SUPPRESSION DURING ACTION OBSERVATION

#### **DISCUSSION**

In this study was investigated whether gaining walking experience by means of a walking training would increase the cortical motor response to observation of walking. The results show that training did not show to have a significant impact on the cortical motor response, as the groups did not differ in their cortical motor response to walking. In both groups, the infants did not display a difference in their response to observing crawling compared to walking movements. These results suggest that there is no difference in motor activation during the observation of an action which is part of the infants' motor repertoire and an action not yet integrated in their motor repertoire.

In the control group, no difference was found in the mu-power while observing crawling compared to walking movements. A similar result was found for the beta-power. These results imply that there was no difference in cortical motor activation during the observation of both movements, although the infants had experience with one movement, yet not with the other. These results are in contrast with the study of van Elk et al. (2008) who found reduced mu- and beta-power during the observation of crawling compared to walking movements. First, the contrasting results might be due to the low amount of infants included in the current study. Due to this low sample size, it is difficult to draw reliable conclusions. As people are variable, systematic differences between conditions might only be found in samples containing more individuals than the one currently presented. However, there might also be another explanation for the disparity in the results of these two studies. When looking at the data of van Elk et al. (2008) it is noticeable that infants only start to show a difference in mu-power for crawling compared to walking movements once they have at least six months of crawling experience, as displayed in Figure 9. Infants with less crawling experience did not show this difference in mu-power for crawling compared to walking videos. The infants included in the current study all had less than four months of crawling experience. Thus, it might be concluded that the infants in our study did not have enough crawling experience yet to be able to find a different cortical motor response to the observation of walking and crawling.

Similar to the control group, the training group did not show a difference in mu- and beta-power while observing crawling compared to walking movements. Thus, these infants also did not show a difference in cortical motor activation during the observation of both movements. The lack of a difference between the two groups indicates that the walking training has no significant effect on motor activation during the observation of walking and crawling. A reason for the inability to find a difference between the two groups might be that one week of training was not long enough for an effect on motor activation to arise. As discussed in the previous paragraph, the study of van Elk et al. (2008) suggests that a sufficient amount of experience with an action is needed before effects on motor activation are found.

Nonetheless, it was expected that one week of walking training would be sufficient to find effects on motor activation. In the study of

Paulus et al. (2012) one week of training with a rattle was enough to find mu-power suppression while listening to the sound made by the rattle compared to a sound not made by the rattle but heard during training and a complete novel sound. Why was the walking training in the current study not as effective as the rattle training in Paulus et al.'s (2008) study? A possible explanation might be that in the study of Paulus et al. (2008) the infants heard the exact same sound during test as heard during training. Thus, the test situation was comparable to the training situation, except that the infants did not shake a rattle. In the current study, the test situation was not comparable to the training situation. The input that the infants received during test was completely different from the input during training. During test, the infants had to sit still, the walker was not in sight and the walking videos were recorded with infants walking without a walker. Therefore, it might have been very difficult for the infants to recognize the walking videos as the action which they had practiced during training.



Figure 9. Correlation between crawling experience and difference in mu-power in the study by van Elk et al. (2008). The infants who show less mu-power for crawling compared to walking all have at least six months of crawling experience (contained in red circle). The other infants do not show a difference. Figure from van Elk et al. (2008), p. 812 (Figure 4).

Another difference between the two studies is that in Paulus' et al. (2012) study the trained action was already part of the infants' motor repertoire. The infants were able to shake a rattle prior to the study. During training the infants acquired an action-effect association through active experience which later elicited a motor response upon hearing sounds. In contrast, the infants participating in the current study were not able to walk yet and thus the trained action was not part of the infants' motor repertoire prior to the study. It might be that a training period of one week is sufficient to find effects on motor activation when the involved action is already part of the individual's motor repertoire. Yet, for actions not part of the motor repertoire prior to the study longer training periods might be needed in order to find effects on the cortical motor response. Further research is needed to define appropriate training periods.

Nevertheless, some similarities in the results of the current study and the study of Paulus et al. (2012) were found. Paulus et al. (2012) found a significant correlation between the amount of training

#### VAN ROOIJEN

and mu-power suppression. Infants who had received most training with the rattle showed the highest mu-suppression while listening to the sound of the rattle. Although insignificant, a comparable line seems to emerge in our data (Figure 5 & 8). The infants who spent most time in the walker showed the smallest difference in mupower while observing crawling compared to walking movements. A similar pattern is found for the beta-power. These results might indicate that longer training durations would have altered the motor cortical response to the observation of walking on a group level. More infants need to be tested and included in the study to be able to confirm this pattern.

In conclusion, the current study provides no evidence for the idea that training has a significant impact on the cortical motor response to action observation. We cannot rule out, based on the current findings, that previously found differences might be due to brain maturation. Longer experience with the actions seems necessary as only experienced individuals show a different motor response in reaction to an action which is part of their motor repertoire compared to an action not part of their motor repertoire (Calvo-Merino et al., 2005, 2006; van Elk et al., 2008). Future research using longer training periods might help disentangle whether one's own action experience changes the processing of the observation of these actions. The effect of action experience on action observation might only arise after extensive periods of training.

#### REFERENCES

- Caetano, G., Jousmaki, V., & Hari, R. (2007). Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. Proceedings of the National Academy of Sciences, 104, 9058-9062.
- Calvo-Merino, B., Glaser, D.E., Grèzes, J., Passingham, R.E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. Cerebral Cortex, 15, 1243-1249.
- Calvo-Merino, B., Grèzes, J., Glaser, D.E., Passingham, R.E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. Current Biology, 16, 1905-1910.
- Cross, E.S., Hamilton, A.F.D.C., & Grafton, S.T. (2006). Building a motor simulation de novo: Observation of dance by dancers. NeuroImage, 31, 1257-1267.
- van Elk, M., van Schie, H.T., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. NeuroImage, 43, 808-814.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain, 119, 593-609.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of action: A meta-analysis. Human Brain Mapping, 12, 1-19.
- Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S.,

& Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. Neurobiology, 95, 15061-15065.

- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A.O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. Journal of Cognitive Neuroscience, 17, 282-293.
- Marshall, P.J., Young. T., & Meltzoff, A.N. (2011). Neural correlates of action observation and execution in 14-month-old infants: an event-related EEG desynchronization study. Developmental Science, 14, 474-480.
- Meyer, M., Hunnius, S., van Elk, M., van Ede, F., & Bekkering, H. (2011). Joint action modulates motor system involvement during action observation in 3-year-olds. Experimental Brain Research, 211, 581-592.
- Muthukumaraswamy, S.D., & Johnson, B.W. (2004). Primary motor cortex activation during action observation revealed by wavelet analysis of the EEG. Clinical Neurophysiology, 115, 1760-1766.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience, 2011.
- Paulus, M., Hunnius, S., van Elk, M., & Bekkering, H. (2012). How learning to shake a rattle affects 8-month-old infants' perception of the rattle's sound: Electrophysiological evidence for action-effect binding in infancy. Developmental Cognitive Neuroscience, 2, 90-96.
- Rizzollati, G., & Craighero, L., (2004). The mirror-neuron system. Annual Reviews Neuroscience, 27, 169-192.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. Topics in Cognitive Science, 1, 353-367.
- Southgate, V., Johnson, M.H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. Biology Letters, 5, 769-772.



# Getting to the bottom of processing behinds

### Willemijn van Woerkom Supervisor: dr. Mariska Kret University of Amsterdam

#### ABSTRACT

Objectives: Evolutionary speaking the processing of behinds is predicted to be very important to chimpanzees and humans, analogous to face processing. Two experiments are described to discover whether two established effects in face processing also hold for behinds: an inversion effect (Experiment 1) and attentional advantage (Experiment 2) of behinds of conspecifics over other object categories. Different color conditions are used to investigate the effect of color. Methods and Results: Experiment 1: a matching-to-sample task in color and greyscale, administered to humans (N = 118, 55 females) and chimpanzees (Pan Troglodytes, N = 5, 4 females). Target categories were human/chimpanzee faces, feet or behinds, or cars, presented upright or inverted. Experiment 2: touch screen oddball in crowd detection task in 3 color conditions with human participants (N = 46, 23 females). Target categories were human/chimpanzee faces. Conclusions: Preliminary evidence was found for a behind inversion effect in chimpanzees. Also, higher accuracy in identifying behinds in both human and chimpanzee males point to expertise in identifying behinds in males but not females. The effect of color was ambiguous. The theoretical consequences of these results and proposals for follow-up studies are discussed.

#### **KEYWORDS**

Sexual Swelling, Chimpanzee, Face Inversion Effect, Attention To Body Parts, Identity Recognition

#### **1. INTRODUCTION**

#### 1.1 The Processing of Faces and Inversion Effects

Face recognition plays an incredibly important part in survival for primates, including humans and chimpanzees. The changeable properties of faces, like expression and gaze, display emotion and can be used to predict behavior. The more or less invariant properties of faces, that do not change on small timescales, are used for identification and display physical characteristics of the owner, like sex and age [16]. Their significance commands a special kind of processing to ensure rapid, reliable and highly specialized recognition. Faces are processed holistically, that is, recognition mainly relies on the configuration of the parts rather than on the identification of parts in isolation [44].

Diamond and Carey [9] defined two types of configurational

information: first order relational information (residing in the positions of the parts relative to each other, used for recognizing a face as such), and second order relational information (position of the parts relative to the prototypical configuration, used for identification). Prototypes are formed during prolonged exposure and are a product of experience, while first order processing might to some extent be innate for both humans ([26], but see [36] for an alternative explanation) and chimpanzees ([32]; [7]). The amount of studies that point to the exceptional form of the processing of faces is overwhelming: faces seem to be both more easily detected than other object categories [35] as better able to subsequently hold attention [1]. Threatening faces in particular seem to draw our attention [3]. Evidence for the exceptional status of faces is found not only in healthy adults, but for example also in visual extinction patients [43] and newborns [21]. Morton and Johnson [29] suggested

#### GETTING TO THE BOTTOM OF PROCESSING BEHINDS

that two circuits in the developing child work together in establishing expert-level processing of faces. First there is an attentional bias towards faces, that is presumably subcortical and present at birth. This bias ensures appropriate input for the later developing, holistic, more cortical face processing.

The special processing of faces causes a peculiar effect first described by Yin [44]. He found that the recognition of faces was disproportionally affected by inversion. Recognition of other object classes, such as houses, was also impaired by inversion, yet not quite as much as face recognition. He called this effect the Face Inversion Effect (FIE). Later studies have shown that this effect can also be observed in expert viewers of other categories, like in dog show judges when viewing dog silhouettes [9], and in non-expert viewers in another very common object category that relies on configurational processing, namely body postures ([33] (behavioral study) and [37] (ERP study)). The finding that inversion effects are not exclusively found in processing faces suggests that there might be other visual object categories that are processed in this specialized manner. One such candidate is the category of behinds.

Behinds are similar to body postures and faces on important features: they have a reliable structure across individuals of the same species, they are ubiquitous in the environment of the subjects which ensures a high level of exposure, and correct interpretation of the conveyed information is crucial for the reproductive success of an individual. Unlike some body postures, faces and behinds share the additional feature of symmetry, thought to be an important characteristic both in lower and higher level visual processing [34]. Lastly, where humans are predominantly furless, including (for males most of the) face and behind, the face and behind are some of the few hairless areas on the body of a chimpanzee. This observation indicates that the visibility of the skin might contribute to the signaling function of these areas, possibly by allowing hemoglobin saturation levels of the skin to be visible (also see section 1.3).

Using images of behinds, De Waal and Pokorny [8] managed to show that chimpanzees are capable of forming some sort of gender construct, at least for familiar individuals. They suggest this construct is possible thanks to whole-body knowledge of familiar individuals (group mates) which allows them to link faces to behinds. This finding hints that from an evolutionary perspective the behind is sufficiently important and recognizable to pay attention to (i.e. a comparable degree of recognition is not expected for arms or legs of group mates). The evolutionary importance and physical properties of behinds suggest that the existence of a Behind Inversion Effect (BehindIE) is not at all unlikely.

In chimpanzees and humans some lower level characteristics of visual perception are similar, like the masking of stimuli, acuity, critical flicker frequency (frequency at which flicker is perceived), and even color categorization [24]. There are suggestions that higher order visual processing, too, is very similar in humans and chimpanzees. Where faces draw the attention of human observers [40], upright faces seem to have a detection advantage over tilted faces in a visual search task done with chimpanzees, an effect that is possibly mediated by an attentional bias. Recent Positron Emission Tomography studies have suggested that chimpanzees even have brain areas that selectively respond to faces [31], comparable to the human fusiform face area [25]. Just like in humans, there is evidence for a FIE in chimpanzees when recognizing chimpanzee and human faces, but not capuchin monkey faces or automobiles [30]. As can be expected given the existence of inversion effects, just like humans, chimpanzees appear to use second order configurational information for face recognition in conspecifics [32]. There have not been conclusive results on whether they also use this type of processing for human faces: Parr et al. [32] found they do not, but see Tomonaga [41] and Dahl et al. [7]. There is an important difference between the chimpanzees used by these two research groups: whereas at the Yerkes National Primate Research Center, chimpanzees are usually approached by humans who wear medical masks, the chimpanzees at the Kyoto University Primate Research Institute often see human faces without masks. Hence, whether chimpanzees do or do not process faces holistically, may depend on the extent of their exposure to human faces, compared to their exposure to the faces of conspecifics.

#### 1.2 Brain Areas Involved in Face Processing

Several brain areas have consistently been linked to the processing of faces; according to Haxby et al. [16] the Fusiform Face Area (FFA), the posterior Superior Temporal Sulcus (STS) and lateral inferior occipital gyri comprise the `core system', charged with the visual processing of faces. The areas differ in their function: some process invariant aspects of faces (including the FFA), and thus facilitate identification, while others are more concerned with changeable properties like direction of gaze (including the STS) [17]. These very different types of information seem to follow a different processing pathway, together forming an integrated system for face processing. The `core system' communicates among itself, but it is also in close contact with the `extended system' that involves the amygdala, insula and limbic system for emotional interpretation, the anterior temporal lobe for retrieving biographical information about the individual, and the intraparietal sulcus, which is involved in spatially directed attention.

The FFA is an area that is tightly involved in inversion effects. It is thought to not only be active when processing faces, but any object class the individual is highly familiar with. Familiarity leads to viewing members of a category as individuals, rather than anonymous exemplars of that category [14]. Since De Waal and Pokorny [8] have shown that chimpanzee are able to identify group members by their behinds, this is strong evidence behind processing in chimpanzees is very advanced and may show inversion effects.

#### 1.3 Color Perception and its Evolutionary Significance

Many species of animals have tri-(or more) chromatic color perception. Color perception is mediated by the existence of different types of pigments in receptor cells in the retina, called cones. These pigments are sensitive to light in a certain wavelength interval. The common ancestor of all terrestrial vertebrates probably possessed

#### VAN WOERKOM

four cone pigments. But two types of spectral cones were lost to most placental mammals when their ancestors became nocturnal. Primates are the only mammals that (re)developed trichromatic color vision [42]. Currently some primate species have uniform/ routine trichromacy, meaning both males and females possess it, and others exhibit polymorphism. This occurs when the gene coding for M and L cones resides on the X-chromosome, resulting in trichromacy only in heterozygous females.

Several different factors are suspected to have played a role in this singular development. It is thought that trichromacy posed an advantage over dichromacy during foraging on colorful fruits and/or certain types of leaves [42], although this conclusion is somewhat debated [18]. An alternative explanation is suggested by the observation that only primates appear to possess the neural wiring to effectively process trichromatic perception [42]. Additionally, the mutation duplicating the gene for the M/L cone might be relatively rare. There appear to be a number of downsides to uniform trichromacy as well. An example is that chromatic aberration in the lens causes light of different frequencies to be broken differently. As such, it is impossible for light of different wavelength intervals, corresponding to different cone types, to be focused on the retina simultaneously when their intervals do not show sufficient overlap. This can seriously interfere with acuity [42]. Lastly, dichromatic vision might in some cases be an advantage in spotting camouflaged prey or predators [38].

Once developed, color vision (and especially trichromatic color perception) proceeded to impose a selective pressure on certain external traits. For instance, the spectral sensitivity of the M and L cones in humans is near optimal for discriminating both density of hemoglobin (the molecule carrying oxygen in blood) and oxygen saturation of the blood [4]. This can be of importance when viewing conspecifics. But this ability is only useful on bare skin, where variations in these dimensions are visible. This 'preexisting bias' [11] could be a factor in explaining the pervasive baldness of the human body and selective absence of hair on chimpanzee faces and behinds [4]. Together with trichromatic color vision, gregarious mating systems (multimale-multifemale, polygynous) also seem to play a role in the emergence of red pelage and skin evolution [11]. It could even be speculated that color vision has played a role in the emergence of gregariousness [11], and has shifted the mode of sociosexual communication away from scent to vision [22].

#### 1.4 Signaling Function of Sexual Swellings

Vision is also important in a very distinct form of sociosexual communication, sexual swellings, which are found in some primate species around the time of ovulation. They are found most often in species with multimale-multifemale mating systems. In these species, the female anogenital region (and sometimes rump) displays swellings and/or coloration for a certain period of time in their menstrual cycle. The swellings are not always reliable predictors of ovulation, and several hypotheses on the exact function of this sexual skin exist. They might be an indication of fertility, age and progress of the estrous cycle that aids male mate selection. Sexual

swellings are very attractive for males and might ignite male-male competition, offering females the change to select the most desirable male. When females mate with several partners, swellings might encourage sperm competition. Besides in multimale-multifemale mating systems, sexual swellings are also found in some primate species that have a polygynous mating system. This suggests that swellings might also mediate female-female competition in obtaining the attention from the one male that is present [10]. The information conveyed by the particular shape of the swelling in individuals is still under investigation [19].

Whatever their exact function might be, the swellings are thought to provide a substantial advantage, because they can be costly: swellings increase female body weight and thus increase energy expenditure [23], their color might make females more visible to predators, they mostly consist of water and thus can be hard to produce in dry environments, and lastly they increase the risk of infections through cuts and fouling by the female's own feces. However, it must be noted that none of these factors have been proven to be life threatening [10].

Although humans, like chimpanzees, have bare behinds without much hair growth, the cyclic swelling and coloration are not (or no longer) present in humans, or at least not that visibly. When looking at the evolutionary tree of New and Old World primates, it becomes apparent that mostly Old World primates display sexual swellings. One possible factor in explaining this difference is the fact that New World primates make more use of odors in communication, and Old World primates rely more on sight. But the direction of a possible causal link between the role of the two modalities is unclear (see section Color perception and its Evolutionary Significance). A related factor might be the presence or absence of color vision, since the reddish or pinkish hue of the sexual swelling is part of the information it communicates. It has also been suggested that in some primate species the extreme swelling and coloration has gradually decreased to prevent recognition from large distances, to decrease the risk of hybridization of species [10].

Although humans generally do not live in a multi-male multifemale society and do not show obvious sexual swellings or colorations, the behind might still have a signaling function. The human female behind still acquires rather large quantities of adipose tissue. This accumulation of fat might have been a sign of fitness in harsh savanna conditions. Besides accumulating adipose tissue on their behinds, human females also develop relatively large breasts. Although some other primates develop pendulous breasts, humans are the only primates to develop breasts before their first pregnancy [10]. Interestingly, breasts may have evolved to resemble the bottom, being more visible when walking upright [28]. Combined, possessing large breasts and behinds might have posed a way of storing energy for females, helping them fulfill the nutritional needs of themselves and their child during pregnancy and nursing. Concentrated areas of fat are more adaptive than a distributed layer over the entire body, because this would significantly alter thermal regulation [10]. In forming a cue of female reproductive fitness, breasts and buttocks may have become objects of sexual attraction. It appears that the waist to hip ratio (WHR) is important in this respect as well, as men from very diverse cultural backgrounds consistently rate women with a lower WHR (0.7) as more attractive than those with a higher WHR (0.9) [10]. The influence of body weight on attractiveness is sometimes thought to differ somewhat across cultures, contrasting the Western, skinny ideal against cultures where large bodies are viewed more positively like certain African and South Pacific cultures. However, these differences might be slowly diminishing due to globalization [39].

#### 1.5 Hypotheses

Summarizing our review, it can be concluded that faces occupy a special place in visual processing, and behinds share many relevant features with faces. As of yet it has not been investigated whether this similarity extends to the existence of inversion effects in observing behinds, in neither chimpanzees nor humans. Inversion effects are typically observed for object categories the viewer has had a high level of exposure with. Given the high level of exposure to the behinds of their group mates, this requirement for inversion effects is met for chimpanzees, but not for humans in our present day society. Additionally, it is thought that attention is likely to play a part in the development of an expertise in viewing faces, since both humans and chimpanzees display an attentional preference to faces over other subject categories. It is unclear whether such a bias also exists for behinds. Since sexual swellings occur in communities with different male-female compositions, there is little reason to assume that males and females would process them differently. Lastly, the emergence of genes coding for trichromatic color vision has had a great impact on the development of the common ancestor of humans and chimpanzees. One of the effects was the employment of color as a medium of sociosexual communication. The expansion of areas on the body showing bare skin is one potential way of utilizing this medium. Given that human and chimpanzee behinds are both furless, the presence of color (red) may mediate the potential attention and inversion effects associated with behinds. Until now this has been an open question.

To answer the open questions described above, we conducted two experiments. In Experiment 1 we focus on the expertise in classifying certain body parts, in Experiment 2, we measure how much subjects attention is drawn towards these cues.

In Experiment 1, humans and chimpanzees conducted a matching to sample task with upright and inverted faces, feet and behinds from unfamiliar humans and chimpanzees. Two versions of this task were prepared; Experiment 1a consisted of colored stimuli, Experiment 1b of greyscale images. Our hypotheses are as follows:

#### Experiment 1a and 1b

Hypothesis 1: Conspecifics' behinds are subject to inversion effects in chimpanzee, but not human, observers.

Hypothesis 2: The presence of color, primarily red, enhances inversion effects for behinds in chimpanzees.

Hypothesis 3: There is no gender difference in humans or chimpanzees in the inversion effect in viewing the behinds of conspecifics. (combined with Experiment 2)

In Experiment 2, humans and chimpanzees will conduct a visual search task with photographs of cars and human and chimpanzee faces and behinds. The `redness' of the photographs was manipulated.

#### Experiment 2

Hypothesis 4: Chimpanzees and humans share an attentional bias towards the behinds of conspecifics over other object categories.

Hypothesis 5: This attentional bias is comparable to that of faces.

Hypothesis 6: The presence of color, primarily red, enhances attention effects for behinds in chimpanzees.

#### 2. EXPERIMENT 1

#### 2.1 Method

Experiment 1 was split up into two sub-experiments, that shared the same procedure but differed slightly in the type and color of the stimuli that were used. These sub-experiments will be denoted as Experiment 1a and Experiment 1b.

#### 2.1 Experiment 1a

#### 2.2.1 Participants

Chimpanzees (Pan Troglodytes: N = 5, 4 females (mean age = 29.0 yrs., std. = 9.9 yrs. (range 12-36 yrs.), 1 male (age = 12 yrs.) from the Primate Research Institute in Kyoto, Japan (for additional information on the chimpanzee participants, see Figure 23 in Appendix A).

Humans: N = 61, 30 females (mean age = 20.8 yrs., std. = 3.3 yrs., range = 18-28 yrs.), 28 males (mean age = 24.0 yrs., std. = 4.9 yrs., range = 19-41 yrs.), 3 participants with unrecorded gender (mean age = 22.2 yrs., std. = 4.7 yrs., range = 19-29 yrs.). All human participants (including those in Experiment 1b, Experiment 2 and the validation study) were heterosexual.

	Experiment 1	Experiment 2	
Human Faces	HFa		
Human Behinds	HBe1	HBe2	
Human Feet	HFe -		
Chimpanzee Faces	CFa		
Chimpanzee Behinds	CBe		
Cars	Cars1	Cars2	

Figure 1. Photograph categories used in the two experiments, and their abbreviations. Note that there are two different categories for human behinds and cars.

#### 2.2.2 Stimuli

The stimuli photographs were all taken under well lit conditions without flash. For the current experiment there were six stimulus categories: human behinds (HBe1), human faces (HFa), human feet (HFe), chimpanzee behinds (CBe), chimpanzee faces (CFa) and chimpanzee feet (CFe) (for an overview of the different stimulus

categories and their abbreviations , see Figure 1).

The individuals in all the chimpanzee photographs live at the Kumamoto Primates Research Park, Japan, and were obtained from Mori et al. [27]. The pictures of the chimpanzee behinds were taken at maximal tumescence. The photos of human behinds depicted the anogenital regions of three human females, and were made to resemble the chimpanzee behinds as closely as possible. The photos of human and chimpanzee faces were frontal photographs. Upright and inverted photographs from all categories were adapted in Photoshop (to isolate/cut out the body parts) and used as stimuli. Several different photographs of each individual were used. None of the individuals shown in the stimuli were familiar to any of the subjects, and human participants were notified that the persons in the HFa category were not the same as those in the HBe1 category. Chimpanzee and human participants were presented with the same stimuli. All stimuli in Experiment 1a were presented in full color.

#### 2.2.3 Procedure

In Experiment 1a human and chimpanzee subjects performed a matching-to-sample task. The chimpanzees performed the task on a touch screen, while human participants performed the task using a computer screen and a button box.

Each trial consisted of the presentation of a photograph that was presented up to two seconds. Then the photograph disappeared from the screen and immediately two new photographs were presented. One of these depicted an individual with the same identity as the photograph that was displayed before, and the other a stimulus from the same category, but from a different individual. Subjects had to touch on the matching photograph. For instance: participants would see a photo of the face of chimpanzee A. After this, two new photos would be displayed: a different photo of the face of chimpanzee A, and a photo of the face of chimpanzee B. In this case the correct response would be the photo of chimpanzee A. See Figure 2 for a graphic representation of a trial.



Figure 2. Schematic representation of a trial in Experiment 1. The images come from the chimpanzee feet (CFe) category. In this trial the left image is the correct response.

The chimpanzee participants started with eight training sessions that were spread out over two days. Stimuli consisted of Japanese castles (they can see such a castle from their outdoor compound). When performance was at 80% correct, the first session of Experiment 1 was started. All chimpanzees were tested on five separate days, except one male juvenile, whose sessions were more spread out due to external factors. Each day, the chimpanzees completed four sessions: two color sessions from Experiment 1a, composed of 72 trials, and two greyscale sessions from Experiment 1b, composed of 84 trials. In total, all chimpanzee participants completed 720 trials for Experiment 1a and 840 trials for Experiment 1b. Human subjects participated in only one session that consisted of 170 or 180 trials for Experiment 1a, or 168 trials for Experiment 1b. Chimpanzees received a piece of apple after a correct response. In case of an incorrect response, the trial was replayed, showing the correct answer. No action from the chimpanzee was required during this re-play trial. For the human participants, no piece of apple or re-play trial followed a correct or incorrect response, respectively.

#### 2.3 Experiment 1b

#### 2.3.1 Participants

Chimpanzees: Chimpanzee subjects were the same as in Experiment 1a.

Humans: For Experiment 1a, a different group of human subjects was tested than for Experimen t 1b. N = 57, 25 females (mean age = 20.7 yrs., std. = 2.1 yrs., range = 18-25 yrs.), 15 males (mean age = 23.0 yrs., std. = 5.2 yrs., range = 18-37 yrs.), 17 participants with unrecorded gender (mean age = 22.3 yrs., std. = 5.1 yrs., range = 18-35 yrs.).

#### 2.3.2 Stimuli

The stimuli were the same as for Experiment 1a. Additionally, frontal photographs of cars were added as an extra stimulus category (category Cars1 in Figure 1). All stimuli in Experiment 1b were presented in greyscale.

#### 2.3.3 Procedure

Task procedure was identical to the procedure of Experiment 1a and is described in Section 2.2.3.

#### 2.4 Results

Before analyzing the data, outliers were removed. An indispensable step since the chimpanzee participants would occasionally get distracted from the task by things like the cries of group mates from another room, and cease responding for some period of time. Human participant data was subjected to the exact same procedure. First, extreme outliers, that skew the mean substantially, were removed by excluding all trials with a reaction time longer than the median per species + 4000 ms. Given the nature of the task this was deemed a reasonable limit, above which response times were likely to signal abnormal trials. Subsequently trials with a reaction time exceeding the personal average reaction time plus two standard deviations were removed. The data from five participants was excluded from analysis due to a technical error during the task. Three other participants were excluded because they performed at chance level for multiple target stimuli categories, among which human upright faces, a category for which a higher performance

level may be expected if the task is performed correctly. Final human samples were N = 56, 25 females, 28 males, 3 unrecorded gender for Experiment 1a, and N = 54, 25 females, 13 males, 16 unrecorded gender.

The RStudio software package (Version 0.97.551) was used for imaging and exploration of the data. Data analysis was done using IBM SPSS Statistics (Release 21.0.0.0). Mixed effects linear modeling was used. This method is more advanced than repeated measures ANOVA (rANOVA), a common method for analyzing repeated measures data. Contrary to rANOVA, for mixed models the assumption of sphericity does not have to hold [2]. Some other advantages of mixed models over rANOVA are that trends in individual subjects can be estimated, the model can handle randomly missing observations, and it uses all available data on the subjects (instead of the averaging used in rANOVA). Also, equal spacing between measurements on different time points is not necessary [15]. Generalized mixed models are suitable for both continuous and nominal dependent variables, so the RT and ACC data in the current experiment can be analyzed using the same method [13]. But perhaps most importantly, this method can take into account the fact that repeated observations of the same individual are often highly correlated, so it does not assume independence between cases by allowing explicit modelling of the nesting of repeated observations within subjects. If not corrected for, the dependence between cases can case a dramatically inflated chance of Type I errors [12].

Analysis of Experiment 1 is not yet completed; for the preliminary results as presented in this report, cars will not be included in the results. This leaves us with the independent variables participant species and gender, and stimulus species (human/chimpanzee), body part (face/foot/genital) and color (color/greyscale). These latter four variables will be referred to as gender, stimspecies, body part and color, respectively. The dependent variables are reaction time (RT) and accuracy (ACC, whether the correct image was selected or not). The following sections will discuss the results from the RT analysis for the chimpanzee and human subjects separately. After this, ACC data analysis will be described, again seperately per participant species.

#### 2.4.1 Human Reaction Time Analysis

The distribution of the dependent variable RT significantly deviated from the normal distribution : skewness (statistic = 1.636, std. error = 0.018) and kurtosis (statistic = 4.193, std. error = 0.035) were both significant. Therefore a gamma target distribution was used, since that is meant for dependent variables like RT that are skewed to the right and have only positive values. The possibility of using non-normal target distributions is one of the advantages of generalized mixed model analysis over general mixed models and rANOVA [13].

In the final model, variance in intercepts over subjects was significant (var( $\mu_{oJ}$ )=0:043,  $\chi^2(1)$  = 6:110, p =0:225). Gender is nearly significant (F (1, 10483) = 3:338, p =0:068); females show a slightly lower reaction time. Color (F (1, 10483) = 3:927, p =0:048) and body part (F (2, 10483) = 87:601, p< 0:000) were both significant, as was their interaction color\*body part (F (2, 10483) = 14:915, p<

0:000). Human participants responded faster to face and foot trials than to trials with behinds, and faster to colored trials than greyscale trials. The difference between colored and greyscale trials was significantly larger for behinds than for faces and feet (Figure 3).





The main effect of species (F (1, 10483) = 19:717, p< 0:000) was significant too, as was its interaction with body part (F (2, 10483) = 176:09, p< 0:000). RT for human stimuli was slightly lower than for chimpanzee stimuli, but this effect depended strongly on body part: RT for chimpanzee feet was lower than for human feet, while it was higher for chimpanzee faces than for human faces, and almost the same for human and chimpanzee behinds (see Figure 4). Finally the 3-way interaction species\*body part\*orientation (F (2, 10483) = 7:373, p =0:001) was significant, signalling a significant difference between upright and inverted human faces, but not for any other human or chimpanzee body part (see Figure 4).



Figure 4. Human participants RT analysis, species\*body part\*orientation.

#### 2.4.2 Chimpanzee Reaction Time Analysis

In the final model, variance in intercepts over subjects (var( $\mu_{oJ}$ )=0:020,  $\chi^2$  (1) = 1:213, p =0:225) nor variance in intercepts subject\*session were significant (var( $\mu_{oJ}$ )=0:000,  $\chi^2$ (1) = 1:623, p =0:105). A main effect of color was found (F (1, 5000) = 11:335, p =0:001); RT for colored trials was slightly lower than for greyscale trials. The RT for behinds was lower than the RT for faces or feet as expressed in the significance of body part (F (2, 5000) = 49:025, p< 0:000). The third significant main effect was species (F (1, 5000) = 34:915, p< 0:000), since the chimpanzee participants responded faster to human trials than trials depicting conspecifics. This last effect is further divided in

#### VAN WOERKOM

the interaction color\*species (F (1, 5000) = 21:423, p< 0:000), which indicates that the RT was lower for human color trials compared to human greyscale trials, while there was no difference between the color conditions for the chimpanzee trials. This accelerating effect color appears to have on RT is stronger for behinds and feet than for faces (color\*body part (F (2, 5000) = 6:941, p =0:001), Figure 5).



Figure 5. Chimpanzee participants RT analysis, color\*body part.

The significant interaction gender\*body part (F (2, 5000) = 21:465, p< 0:000) appears to be a consequence of the presence of only one chimpanzee male participant, and resulting large error bars. The interaction body part\*orientation (F (2, 5000) = 3:891, p = 0:020) is more interesting, since the inverted condition elicited longer RTs in the face and behind condition compared to the upright condition, but not for the foot condition (see Figure 6).



Figure 6. Chimpanzee participants RT analysis, body part\*orientation.

#### 2.4.3 Human Accuracy Analysis

In the final human accuracy model, variance in intercepts over subjects was significant (var( $\mu_{oJ}$ )=0:193,  $\chi^2(1)$  = 4:806, p< 0:000).

The main effects of orientation (F (1, 12341) = 12:665, p < 0:000) and species (F (1, 12341) = 285:16, p < 0:000) were both significant. As can be expected, accuracy was higher for upright images compared to inverted images, and higher for human than chimpanzee images.

The main effects of gender (F (1, 12341) = 1:242, p =0:265) and color (F (1, 12341) = 0:211, p =0:646) were both non-significant, contrary to the interaction gender\*color\*species (F (1, 12341) = 6:856, p =0:009). Males scored slightly higher on accuracy for greyscale human trials than females, but not for colored human trials. Also females' accuracy was slightly lower than males for

colored chimpanzee trials, but they scored the same on greyscale chimpanzee trials (see Figure 7).



Figure 7. Human participants ACC analysis, gender\*color\*species.

The interaction gender\*body part (F (2, 12341) = 3:524, p =0:029) was also significant, indicating that females and males obtained equal scores for faces and feet, but males scored significantly higher for behinds (see Figure 8).



Figure 8. Human participants ACC analysis, gender\*body part.

The main effect of body part (F (2, 12341) = 2:207, p =0:11) did not reach significance, while its interaction species\*body part (F (2, 12341) = 111:358, p< 0:000) was highly significant. For the accuracy on trials depicting feet, it did not matter whether they belonged to a human or chimpanzee. For faces and behinds, however, participants scored higher for human trials than chimpanzee trials (see Figure 9).



Figure 9. Human participants ACC analysis, species\*body part.



For the chimpanzee accuracy model, variance in intercepts over

subjects was not significant (var( $\mu_{_{0J}}$ )=0:063,  $\chi^2$  (1) = 1:148, p =0:251).

Since Experiment 1 was aimed at uncovering a behind inversion effect, orientation is the most relevant independent variable, but unfortunately it was not significant as main effect (F (1, 6910) = 0:410, p =0:522), nor in any of its interactions. No main effect of gender was found (F (1, 6910) = 0:410, p =0:522) but its interactions gender\*species (F (1, 6910) = 9:462, p =0:002) and gender\*body part (F (2, 6910) = 6:942, p =0:001) were significant (see Figure 10). The male chimpanzee scored higher than the females on chimpanzee stimuli, but they scored similarly on human stimuli. The male also scored higher than the females on faces and behinds, but not feet.





Additionally color (F (1, 6910) = 41:456, p< 0:000), species (F (1, 6910) = 10:050, p = 0:002) and body part (F (2, 6910) = 12:073, p< 0:000) were significant, as were all of their interactions: color\*species (F (1, 6910) = 10:591, p =0:001), color\*body part (F (2, 6910) = 10:651, p< 0:000), species\*body part (F (2, 6910) = 24:994, p< 0:000), color\*species\*body part (F (2, 6910) = 13:389, p< 0:000). The chimpanzee participants scored higher on accuracy in the color condition compared to the greyscale condition for the chimpanzee face and chimpanzee behinds categories, but not for chimpanzee feet (see Figure 11). For all body parts, the chimpanzee participants scored better in the color category than in the greyscale condition, but this effect was most pronounced for behinds (see Figure 12).



Figure 11. Chimpanzee participants ACC analysis, color\*species\*body part.

In the color condition, the chimpanzee participants scored significantly higher for the human behinds than for the chimpanzee behinds. This difference did not occur between human and chimpanzee behinds in the greyscale condition.



Figure 12. Chimpanzee participants ACC analysis, color\*body part.

#### 2.5 Conclusion

Given the modest progress that has been made on the analysis of Experiment 1, only cautious conclusions can be drawn. However, the diversity of both human and chimpanzee data, and the use of both RT and ACC data allows for the description of a number of detectable trends.

A clear effect of color is observed. Both humans and chimpanzees respond faster to color than to greyscale images, mostly so for behinds. Chimpanzees respond faster to colored behinds and feet, but not faces, compared to RT of their greyscale counterpart. Chimpanzees responded more accurately to human and chimpanzee behinds, and chimpanzee faces. Some caution must be observed since the darkness of the chimpanzee faces and feet may have decreased the difference in these two categories between the color and greyscale condition. Chimpanzee behinds, human faces and human feet are all pink. For a more thorough discussion of the variable color, see Section 3.2 and 6.2.

The main focus of Experiment 1 was the inversion effect: lower RT and higher ACC for the upright condition compared to the inverted condition. Some evidence for the face inversion effect was found, since human participants responded faster to human upright faces than inverted faces. Chimpanzee participants also responded faster to faces in the upright versus the inverted condition, and, importantly, showed a similar trend for behinds.

The effect of gender is not equivocal (see for instance the complex interaction gender\*color\*species in Figure 7), but some effects look decidedly interesting: human males score higher on behinds than females (Figure 9), an effect also seen in chimpanzee participant s (Figure 11). This may not point to an inversion effect since orientation is not involved, but it does point to an increased proficiency in males in identifying behinds than females.

#### 3. EXPERIMENT 2



#### 3.1.1 Participants

Experiment 2 was conducted with a different set of participants from Experiment 1. Participants were all human, although as noted before it would be very informative to administer this task to chimpanzees as well. Participants (N = 46, 23 females (mean age = 21.9 yrs., std. = 3.8 yrs., range = 18-35 yrs.) and 23 males (mean age = 26.5yrs., std. = 5.3 yrs., range = 18-41 yrs.) were tested at the University of Amsterdam (UvA). Participants were attracted via social media and through the website for psychological research of the UvA. Participants were given a monetary reward for participation.

#### 3.1.2 Stimuli

Stimuli consisted of six circular images, presented in two rows of three: one target stimulus, and five distractor images (see Figure 13). The target stimulus categories, also found in Table 1, were human faces (HFa), human behinds (HBe2), chimpanzee faces (CFa) and chimpanzee behinds (CBe). The distractor images were five frontal photographs of cars (Cars2). Three versions of the stimuli were created: full greyscale, medium greyscale and color.

These subject categories were chosen for several reasons. Human faces were included in this study to reproduce the attentional bias towards faces as reported in the literature. Human behinds were the objects of interest, since uncovering an attentional bias towards them was the main aim of this experiment. Chimpanzee behinds were included to provide a control category for the human behinds. They resemble human behinds in important respects; hue, being symmetrical, and consisting of biological tissue. They differ from human behind in one important feature; shape, making them a perfect control category. Additionally, they allowed us to balance the design for human and anticipated future chimpanzee participants by including chimpanzee faces as well, adding the independent variable of species to our analysis. Results are expected to be stronger for conspecifics than for the other species.

Apart from method of editing and presentation, the HFa, CFa and CBe images were identical to the ones in Experiment 1 (see Table 1). However, different photos of human behinds were used. Subjects were photographed while standing upright in three slightly differing foot configurations. They showed only the behinds of the individuals, compared to the entire anogenital region for the HBe1 condition. All photographs were taken without flash under well lit conditions. The subjects in the photographs signed an informed consent form. All three were female, in their reproductive age and of a healthy weight (details can be found in Figure 24 in Appendix B).

All photographs were first cut out using Photoshop. For all categories (HFa, HBe1, CFa and CBe) average RGB values were determined, and RGB-values of all photographs in that category were set on these values. Additionally, for all stimulus categories a differently colored distractor `crowd' of cars was created. This was done by giving a set of nine car photographs (Cars2) the average RGB values of the HFa, HBe1, CFa and CBe images, respectively. Additionally, a circle with these same RGB values was placed behind each image. Stimuli were formed using one target image and five distractor images with matching RGB values (see Figure 13 for three examples of stimuli from the CBe category). This procedure ensured that for each stimulus, distractor and target images had the exact same luminosity and RGB values, such that any difference in RT could not be due to differences in color or luminosity.

#### 3.1.3 Procedure

All participants were briefed prior to the task and filled out an informed consent form. Participants were seated approximately 40 cm from the touch screen so it was comfortably within reach, and were instructed to use only their dominant hand for the duration of the task. The dimensions of the touch screen were 47.5 cm by 27 cm, of which the stimulus images took up 20 by 12.5 cm (see Figure 14 for a representation of the configuration of the touch screen during the task). The touch screen occupied approximately 29° by 18° of the visual field of the participants.

All participants completed one session consisting of 216 trials. To ensure standardized hand position, each trial began with the presentation of a green button at the bottom. Participant s were required to press this green button in order to proceed to the stimulus, which was immediately presented for 500 ms. Participants were required to tap the target image as quickly as possible. To increase the difficulty level of the task, after 500 ms the stimulus was replaced by an image of six grey circles (see Figure 15 for a graphic representation of one trial). Note that after the transition from images to grey circles, a response was still required. The participant's choice was followed by a variable intertrial interval, with values between 3000 and 3500 ms in order to avoid spill-over of the effect of one trial to the next. Total testing time for this task, subjects



(a) Stimulus in black and white.

(b) Stimulus in half color.

Figure 13. Examples of stimuli used in the task in the three color conditions. All target images in these examples are chimpanzee behinds (CBe).

(c) Stimulus in full color.

57

#### GETTING TO THE BOTTOM OF PROCESSING BEHINDS

participated in several other tasks that were unrelated to the current study.



Figure 14. Measurements of the touch screen that was used for the task. The locations of the stimulus images are marked by a red dashed line, and the location of the green button is marked by a green dashed line.



Figure 15. Schematic representation of a trial in Experiment 2.

#### 3.2 Results

One female participant did not follow task instructions. Another female participant experienced technical failure during the task. It was decided to exclude both participants' data from further analysis, resulting in a final sample of N = 44, 21 females.

The same data analysis software and method was used for Experiment 1 and Experiment 2. Before starting the analysis, artificially high reaction times brought about by the occasional malfunctioning of the touch screen, as reported by participants during exit interviews, needed to be removed. This was done in the same way it was done for Experiment 1. See Section 2.4 for details.

Independent variables were the participant gender, stimulus species (human or chimpanzee) and body part (face or behind) depicted in the target image, and the color of the stimulus (greyscale, half color or full color). These variables will be referred to as species, body part and color, respectively. Reaction time (RT) was the dependent variable. In addition, attractiveness scores per participant per image were added. These were obtained from the participants who also took part in the validation experiment, and formed the independent variable attractiveness.

The distribution of the dependent variable RT significantly

deviated from the normal distribution: skewness (statistic = 2.454, std. error = 0.026) and kurtosis (statistic = 8.186, std. error = 0.051) were both significant, as was the Kolmogorov-Smirnov test (statistic = 0.221, df = 9054, sig. = 0.000). As with the Experiment 1 RT-analysis (see Section 2.4.1) generalized linear mixed models and a gamma target distribution were used.

For the first models, three of the significant fixed effects were body part, species and body part\*species. However, these effects were driven mainly by the significantly higher RT of chimpanzee face trials compared with human faces and behinds, and chimpanzee faces (see Figure 16). It was suspected that this difference may be caused by the low luminosity of the chimpanzee face images. The mean luminosity of the chimpanzee face target images was 108 (std. = 1.6) candela/m<sup>2</sup>, while the luminosity values for the CBe, HFa and HBe2 were 165.6 (std. = 1.8) candela/m<sup>2</sup>, 148.6 (std. = 1.9) candela/m2 and 150.8 (std. = 1.6) candela/m<sup>2</sup>, respectively. This unwanted effect caused us to exclude the chimpanzee face categories from further analysis. The expected consequences of this decision will be discussed in Section 6.2.





A new independent variable, condition, was introduced to replace body part and species, since their main effect could not be investigated anymore with the elimination of the CFa category. The three categories of condition are chimpanzee behinds, human faces and human behinds.

In the final model, variance in intercepts over subjects was significant (var( $\mu_{oJ}$ )=0:024,  $\chi^2(1)$  = 4:534, p<0:000). Gender was not significant, nor as main effect nor in any interaction . Attractiveness was also not a significant predictor of RT. Condition (F (2, 6779) = 3:686, p =0:025) and color (F (2, 6779) = 8:174, p< 0:000) on the other hand, were both significant, as was their interaction (F (4, 6779) = 5:686, p< 0:000). Human behinds elicited a significantly longer RT than chimpanzee behinds, and participants responded significantly faster to the half color condition than to the full color condition. At least part of these effects, however, were driven by the remarkably fast reaction times for the CBe category in the half color condition (see Figure 17). Furthermore, RT to human faces was significantly slower for the full color condition than the greyscale and half color condition.



Figure 17. Human participants RT analysis, condition\*color.

#### 3.3 Conclusion

The significant effects in our analysis were not predicted by our initial theoretical framework. The longer RT for the chimpanzee faces was unexpected, but can be explained using the mean luminosity of the images in that category. Although administering this exact same task to chimpanzee participants may have to be reconsidered, omitting the chimpanzee faces from this analysis is no immediate problem since the chimpanzee behinds still function as a control category for the human behinds.

The markedly low RT for the chimpanzee behinds in the half color condition compared to the other two color conditions is puzzling, because the half color condition was explicitly designed to be in between the full color and greyscale condition. Any effects were expected to be largest in any one of the two extreme conditions, not in the middle one. The relevance of this finding is discussed in more detail in Section 6.

#### 4. VALIDATION

Some of the images used in Experiments 1 and 2 can be considered rather explicit, particularly the images in category HBe1 and HBe2. For this reason it was decided to perform a validation experiment, to rate all the target images on attractiveness. This allowed us to check the attractiveness scores for deviant values, explore the relationship between color and attractiveness, and check whether the attractiveness of an image alone would be able to explain some of the variance in the data. Additionally the effect of color on attractiveness could be investigated.

#### 4.1 Method

#### 4.1.1 Participants

All participants were human (N = 42, 21 females and 21 males). All but three participated in Experiment 2 as well, and performed this validation task after completing Experiment 2.

#### 4.1.2 Stimuli

All target face and behind images from Experiment 1 and Experiment 2 were included in this validation task, the categories HFa, HBe1, HBe2, CFa and CBe. All participants rated 135 images, shown in random order: nine images from each condition, in the three color conditions of Experiment 2 (greyscale, half color, full color). The

images were presented in the format of Experiment 2: included in a circle of the average RGB color of the category (for an example of these circles, see Figure 15). This was done to avoid effects of presentation method.

#### 4.1.3 Procedure

Participants were seated in front of a regular computer screen on which the stimuli were shown. Task instructions were: "How attractive do you find this image? Please rate it on a scale from 1 to 5, 1 representing `not attractive at all', and 5 representing `very attractive'." Images were shown until a response was registered. The attractiveness scores per image per color condition from the participants that also participated in Experiment 2 were added to their data (see Section 3.2).

#### 4.2 Results

Two participants displayed a deviant response pattern, rating the images with only two numbers where all other participants used a minimum of three. It was suspected they did not rate the images seriously and they were excluded from validation results analysis, resulting in a sample of N = 40 (20 females, 20 males). The same data analysis software and method was used as for Experiment 1 and 2. Skewness (0.579, std. error = 0.033) and kurtosis (-0.890, std. error = 0.067) were both significant for the dependent variable, the attractiveness rating, so again a gamma target distribution was selected.

In the full mixed model, the variance in intercepts over subjects was significant (var( $\mu_{ol}$ )= 0:106,  $\chi^2(1)$  = 4:307, p< 0:000). No main effect of gender was found (F (1, 5390) = 1:372, p =0:242), but the interactions gender\*body part (F (1, 5390) = 142:313, p< 0:000) and gender\*species\*body part (F (1, 5390) = 4:896, p =0:027) were significant. As can be seen in Figure 18, males rate behinds significantly higher than females, but not faces. Males rated both human behinds as chimpanzee behinds higher than females, although the difference between the sexes is greatest for human behinds.



Figure 18. Human participants attractiveness rating analysis, gender\*species\*body part.

The main effects of species (F (1, 5390) = 1181:115, p< 0:000) and body part (F (1, 5390) = 1228:478, p< 0:000) were both highly significant, as was their interaction (F (1, 5390) = 188:207, p< 0:000). Human images were rated significantly higher than chimpanzee images, and faces were rated significantly higher than

behinds. The difference between human and chimpanzee faces was less great than the difference between the attractiveness rating of human behinds and chimpanzee behinds, the latter of which scored, on average, barely above the minimum score at an estimated mean of 1.28 (std. error = 0.068) (see Figure 19).



Figure 19. Human participants attractiveness rating analysis, species\*body part.

Lastly, the color of the stimulus was almost significant in predicting attractiveness scores at p = 0.064 (F (2, 5390) = 2.745), showing a trend of higher score for greyscale images compared to full color images (see Figure 20).





A follow-up model was built, including only the two categories of human behinds: HBe1 (used in Experiment 1, showing the entire anogenital region) and HBe2 (used in Experiment 2, showing merely the buttocks) to investigate whether "explicitness" of the image (henceforth indicated by condition) had any effect on the attractiveness rating. This turned out to be the case, with both the main effects of gender (F (1, 2148) = 5:654, p =0:018) and condition (F (1, 2148) = 5:572, p =0:018) as their interaction gender\*condition (F (1, 2148) = 59:035, p< 0:000) being significant. The more explicit behinds, HBe1, were rated as being significantly more attractive than the less explicit behinds, HBe2. Males ranked both types of behinds higher than females, but the explicit behinds even more so than the less explicit behinds (see Figure 21).

#### 4.3 Conclusion

From this attractiveness validation study the following conclusions

can be drawn. First of all there is a clear effect of gender; males differ surprisingly little from women in their ratings of human female faces, but significantly in their rating of human behinds. This indicates that the attempt to create stimuli that are more attractive for heterosexual men than women has succeeded. Surprisingly, men even rated chimpanzee behinds higher than women. It could be speculated this may be due to the slight resemblance that still exists between the human and chimpanzee anogenital region.

Apart from this effect of gender, an effect of species was found. The fact that human stimuli were rated higher than the chimpanzee stimuli may not be surprising, but the interaction species\*body part further substantiates the claim that the stimuli form biologically valid representations of their condition, and thus elicit comparable responses.



Figure 21. Human participants attractiveness rating analysis, gender\*condition.

Lastly a near-significant effect of color was found, with a trend towards higher ratings of more greyscale images compared to the more colorful images.

#### 5. CONCLUSION

In this section, some results from the analyses from the data from Experiment 1, Experiment 2 and the validation will be combined in an attempt to come to overarching conclusions. Ultimately, Section 6 will conclude with remarks on how to successfully improve and extend the current study.

In the validation, a near significant effect of color was found, the trend being that greyscale images were rated as more attractive than color images. This is important for two reasons.

First, it is unclear whether our color manipulation had the desired effect. The most extreme effects or scores were expected in the full color or greyscale condition, not the half color condition (for instance the mystifyingly short half color chimpanzee behind RTs in Experiment 2). This holds for the results from the validation study, but not for the results of Experiment 2.

Second, the effect of color on attractiveness ratings is consistent with the findings of Johns et al. [20], who used a slightly different color manipulation than the current study. Their stimuli ranged from dark red to pink, whereas those in the current study ranged from completely greyscale to full color<sup>1</sup>. But since the anogenital regions of chimpanzee females during full tumescence also show

<sup>1</sup> However, since our target images were all fairly pink in the full color condition, in practice the color ranges may have been at least partly overlapping.

#### VAN WOERKOM

a slight (red) coloration, we hypothesized that color would enhance any inversion effects or attention benefits for behinds. We expected these effect to be correlated with an enhanced attractiveness. This did not come up in our data, since adding attractiveness to the Experiment 2 model did not significantly improve it. Attractiveness is not a significant predictor of reaction time in our task, leading us to conclude attractiveness is not a mediating or correlated factor in any attentional effect associated with behinds (although see Section 6.3 for a discussion on the attractiveness measure that was used).

Linked to the effect of color in Experiment 2 and the validation, is the effect of color in Experiment 1. The positive effect of color (decreased RT, higher ACC) was strongest for behinds. This finding is compatible with the hypothesis concerning color that it enhances any attentional bias that exists towards behinds. Some caution is warranted, however, since the difference between color and greyscale is much larger for the `pink' categories (human faces, human feet, human behinds, chimpanzee behinds) than for the `dark' categories (chimpanzee faces and feet). This could influence both the main effects of color, species and body part as their interaction.

Experiment 2 was the experiment designed to uncover an attentional bias towards behinds. Although the results of Experiment 2 provided no evidence for such a bias, the results of the RT analysis of Experiment 1 lead to a different view. Both human and chimpanzee participants responded faster to colored stimuli, but this effect was strongest for behinds. In other words, adding color to images of behinds elicited faster RTs. An effect of gender can be observed in the ACC scores, since both human and chimpanzee males scored higher on behinds than females.

Experiment 1 was aimed at investigating the amount of expertise in identifying behinds, using an inversion effect. Evidence for the classical face inversion effect was found in humans participants: they responded faster to upright versus inverted human faces, while the orientation did not significantly matter for other stimulus categories. More interestingly, some evidence for not only a face inversion effect but a behind inversion effect was found in chimpanzee participants; they responded faster to upright compared to inverted chimpanzee faces and behinds.

Although time constraints prevented a thorough analysis of the data, altogether the results look promising as males appear to be better at identifying behinds than females, and chimpanzees responded faster to upright compared to inverted faces and behinds. This effect of orientation was not present for feet. At the same time we have taken the first steps to rule out attractiveness as a mediating factor in these effects. The effect of gender on accuracy when classifying behinds appears to exist in both humans and chimpanzees. It would be very informative to see at which point in the development this advantage arises by testing participants in a broader age range. This may also provide clues as to why the advantage exists; do males perhaps have an attentional bias towards behinds, but not females? Is this bias an effect of becoming fertile, or is it present from birth? Additionally, administering Experiment 2 to chimpanzee participants, complimenting the humans that have already been tested, may provide invaluable insights; since Experiment 1 has shown that chimpanzees potentially display a behind inversion effect in terms of reaction time, the chance that they have an attentional bias towards behinds cannot be ignored. Their participation may shed further light on the considerable overlap (and interesting differences!) between humans and chimpanzees when processing behinds.

#### 6. DISCUSSION

The combination of Experiment 1, Experiment 2 and the validation study was aimed at uncovering an attentional bias and/or an inversion effect in the processing of behinds. However promising our results, several elements of procedure, data processing and follow-up studies deserve attention to improve interpretability of the results, and the strength of our conclusions. Therefore this Discussion will address several focus points for the study in general, and subsequently Experiment 1, Experiment 2 and the validation task separately.

The first possible point of improvement is the age distribution of the human and chimpanzee participants. The age ranges of the current participant base did not allow testing for an effect of age on inversion effects, attentional bias or attractiveness scores. This problem is most apparent for our chimpanzee participant group, since it only contained one male. But also for our human participants, the mean age and standard deviation of the male and female populations differed significantly, which means that any effect of age found in this study could theoretically be ascribed to the factor age. Besides the correlation with gender, our age distributions are not representative of the human (or chimpanzee) fertile age range. We would expect inversion and attention effects associated with behind to possibly be stronger for those inside, than for those outside of this range. To be able to include age in our models and investigate these hypotheses we would need to test some additional men and women in order to balance the age distribution between the sexes and over a broader age range.

#### 6.1 Experiment 1

Exit interviews with participants revealed that matching of the car images was facilitated by the logo. This effect may have influenced task difficulty in both the upright and the inverted conditions, although it is not directly clear in which direction.

#### 6.2 Experiment 2

Analysis of the recorded reaction times revealed that the touch screen recorded reaction times following the refresh rate of the monitor (60 Hz). This means that the overwhelming majority of recorded reaction times had -17, -33, -50, -67, 83 or -00 as their last two digits. Unfortunately, this effect decreases the accuracy of the recorded reaction times. However, since all responses were affected equally by this effect, we believe that it does not compromise the conclusions that can be drawn from our results.

The subjects photographed for the HBe2 stimuli are very similar in all measured parameter s (age, height, BMI). Since the subjects are all healthy reproductive females, whose behinds are most important to monitor according to our hypothesis, these stimuli are thought to be effective samples for our current purposes. However, future research might include behinds from a broader range of body types and ages.

For all stimuli a grey (R = G = B = 163 candela/m<sup>2</sup>) background was used. It could be argued that the half and full color images stood out more from the grey background than the greyscale images. To level the contrast between the crowd and target images and the background, it might be better to use a different color background, for instance white, against which all color conditions stand out equally.

In Experiment 2 reaction times for trials with chimpanzee faces as target image were significantly higher than for behinds or human faces. Chimpanzee faces were included in the experiment to balance the design and be able to administer the exact same stimuli to human and chimpanzee participants. Excluding chimpanzee faces from the human participants' data analysis may compromise our ability to compare human and chimpanzee results. Chimpanzee participants have not been tested yet, and it would be interesting to see if for them, too, reaction times are higher for chimpanzee faces than for other categories. The results of Experiment 1 suggest this might be the case; the difference in chimpanzee reaction times between colored and greyscale behinds was larger than the difference between colored and greyscale faces. A possible solution for this issue would be to repeat the task, and to replace the current dark chimpanzee faces with lighter ones that approach the luminosity values of the other `pink' categories. Unfortunately, the owners of lighter chimpanzee faces are usually the young ones (see Figure 22), so that exchange would likely introduce more confounding effects than it would solve.



Figure 22. Light-faced baby chimpanzee with her dark-faced mother in Los Angeles Zoo [5].

Although Experiment 2 was designed to measure an attentional bias towards behinds, the chosen format (touch screen task resembling the face-in-the-crowd paradigm) may not have been the best way of achieving this; after visually processing the scene, the motor system still has to perform the action of moving a finger towards the desired position. Another, more direct proxy for attention is eye gaze, measured by an eye tracker. A preferential looking task using an eye tracker would be a natural follow-up or extension of the current study.

#### 6.3 Validation

According to our attractiveness ratings, the greyscale images were judged to be more attractive than the full color images. Although this replicates earlier findings, it can be doubted whether these ratings represent the kind of attractiveness that is relevant for our study. Participant s may prefer to look at greyscale images in a laboratory setting and rate them higher, while getting more physically aroused by the color images. This is the type of attractiveness that would be relevant for our results. Therefore, in follow up studies the attractiveness of the target images may be better determined by a different measure, like skin conductance (correlated to affective report and ERP potentials in response to affective stimuli [6]) or some other involuntary response measure, rather than conscious and voluntary report.

#### REFERENCES

- M. Bindemann, M. Burton, I. T. C. Hooge, R. Jenkins, and E. H. F. de Haan. Faces retain attention. Psychonomic Bulletin & Review, 12(6):1048-1053, 2005.
- [2] E. Blackwell, C. F. Mendes de Leon, and G. E. Miller. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. Psychosomatic Medicine, 68:870-878, 2006.
- [3] B. P. Bradley, K. Mogg, S. J. Falla, and L. R. Hamilton. Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. Cognition & Emotion, 12(6):737-753, 1998.
- [4] M. A. Changizi, Q. Zhang, and S. Shimojo. Bare skin, blood and the evolution of primate colour vision. Biology Letters, 2:217-221, 2006.
- [5] Baby chimpanzee and mother Gracie from Los Angeles Zoo. p:// www.zooborns.typepad.com/zooborns/chimpanzee/, 2012.
- [6] B. N. Cuthbert, H. T. Schupp, M. M. Bradley, N. Birbaumer, and P. J. Lang. Brain potential s in affective picture processing: Covariation with autonomic arousal and affective report. Biological Psychology, 52:95-111, 2000.
- [7] C. D. Dahl, M. J. Rasch, M. Tomonaga, and I. Adachi. Developmental processes in face perception. Scientific Reports, 3, 2013.
- [8] F. B. M. de Waal and J. J. Pokorny. Faces and behinds: Chimpanzee sex perception. Advanced Science Letters, 1:99-103, 2008.
- [9] R. Diamond and S. Carey. Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General, 115(2):107-117, June 1986.
- [10] A. F. Dixson. Primate Sexuality. Oxford University Press, 1998.

#### VAN WOERKOM

- [11] A. A. Fernandez and M. R. Morris. Sexual selection and trichromatic color vision in primates: Statistical support for the preexisting-bias hypothesis. The American Naturalist, 170(1):10-20, July 2007.
- [12] A. Field. Discovering Statistics Using SPSS. SAGE Publications Limited, third edition, 2009.
- [13] G. D. Garson. Generalized linear models & generalized estimating equations. Statistical Associates Publishing, 2012.
- [14] I. Gaulthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore. Activation of the middle fusiform `face area' increases with expertise in recognizing novel objects. Nature Neuroscience, 2(6):568-573, June 1999.
- [15] R. Gueorguieva and J. H. Krystal. More over anova; progress in analyzing repeatedmeasures data and its reflection on papers published in the archives of general psychiatry. Archives of General Psychiatry, 61(3):310-317, March 2004.
- [16] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. The distributed human neural system for face perception. Trends in Cognitive Sciences, 4(6):223-233, June 2000.
- [17] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. Human neural systems for face recognition and social communication. Society of Biological Psychiatry, 51:59-67, 2002.
- [18] C. Hiramatsu, A. D. Melin, F. Aureli, C. M. Schaffner, M. Vorobyev, and Y. Matsumoto. Importance of achromatic contrast in short-range fruit foraging of primates. PLoS ONE, 3(10):e3356, 2008.
- [19] E. Huchard, J. A. Benavides, J. M. Setchell, M. J. E. Charpentier, A. Alvergne, A. J. King, L. A. Knapp, G. Cowlishaw, and M. Raymond. Studying shape in sexual signals: the case of primate sexual swellings. Behavioral Ecology and Sociobiology, 63(8):1231-1242, June 2009.
- [20] S. E. Johns, L. A. Hargrave, and N. E. Newton-Fisher. Red is not a proxy signal for female genitalia in humans. PLoSONE, 7(4):e34669, 2012.
- [21] M. H. Johnson, S. Dziurawiec, H. Ellis, and J. Morton. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition, 40:1-19, 1991.
- [22] E. R. Liman and H. Innan. Relaxed selective pressure on an essential component of pheromone transduction in primate evolution. Proceedings of the National Academy of Sciences, 100(6):3328-3332, March 2003.
- [23] A. Matsumoto-Oda and R. Oda. Changes in the activity budget of cycling female chimpanzees. American Journal of Primatology, 46:157-166, 1998.
- [24] T. Matsuno and K. Fujita. A comparative psychophysical approach to visual perception in primates. Primates, 50:121-130, 2009.
- [25] G. McCarthy, A. Puce, J. C. Gore, and T. Allison. Face-specific processing in the human fusiform gyrus. Journal of cognitive

Neuroscience, 9(5):605-610, 1997.

- [26] C. J. Mondloch, T. L. Lewis, D. R. Budreau, D. Maurer, J. L. Dannemiller, and B. R. Stephens. Face perception during early infancy. Psychological Science, 10(5):419-422, September 1999.
- [27] Y. Mori, K. Shimoda, H. Kobayashi, I. Hayasaka, and Y. Hamada. Development of the sexual skin with pubertal maturation in female chimpanzees. Primates, 48:97-101, 2007.
- [28] D. Morris. The Naked Ape: A zoologist's study of the human animal. Bantam Books of Canada, 1967.
- [29] J. Morton and M. H. Johnson. Conspec and conlern: A twoprocess theory of infant face recognition. Psychological Review, 98(2):164-181, 1991.
- [30] L. A. Parr, T. Dove, and W. D. Hopkins. Why faces may be special: Evidence of the inversion effect in chimpanzees. Journal of Cognitive Neuroscience, 10(5):615-622, 1998.
- [31] L. A. Parr, E. Hecht, S. K. Barks, T. M. Preuss, and J. R. Votaw. Face processing in the chimpanzee brain. Current Biology, 19(1):50-53, January 2009.
- [32] L. A. Parr, M. Heintz, and U. Akamagwuna. Three studies on configural face processing by chimpanzees. Brain and Cognition, 62(1):30-42, October 2006.
- [33] C. L. Reed, V. E. Stone, S. Bozova, and J. Tanaka. The bodyinversion effect. Psychologica I Science, 14(4), July 2003.
- [34] G. Rhodes, M. Peters, K. Lee, M. Concetta Morrone, and D. Burr. Higher-level mechanism s detect facial symmetry. Proceedings of the Royal Society B: Biological Sciences, 272:1379-1384, 2005.
- [35] T. Ro, C. Russell, and N. Lavie. Changing faces: A detection advantage in the flicker paradigm. Psychological Science, 12(1):94-99, January 2001.
- [36] F. Simion, V. Macchi Cassia, and C. Turati. The origins of face perception: Specific versus non-specific mechanisms. Infant and Child Development, 10:59-65, 2001.
- [37] J. J. Stekelenburg and B. de Gelder. The neural correlates of perceiving human bodies: An ERP study on the body-inversion effect. Neuroreport, 15(5):777-780, April 2004.
- [38] A. K. Surridge, D. Osorio, and N. I. Mundy. Evolution and selection of trichromatic vision in primates. Trends in Ecology and Evolution, 18(4):198-205, April 2003.
- [39] V. Swami, D. Knight, M. J. Tovee, P. Davies, and A. Furnham. Preferences for female body size in britain and the south pacific. Body Image, 4:219-223, 2007.
- [40] J. Theeuwes and S. van der Stigchel. Faces capture attention: Evidence from inhibition of return. Visual Cognition, 13(6):657-665, 2006.
- [41] M. Tomonaga. Visual search for orientation of faces by a chimpanzee (pan troglodytes): face-specific upright superiority

and the role of facial configural properties. Primates, 48:1-12, 2007.

- [42] M. Vorobyev. Ecology and evolution of primate colour vision. Clinical and Experimental Optometry, 87(4-5):230-238, 2004.
- [43] P. Vuilleumier. Faces call for attention: Evidence from patients with visual extinction. Neuropsychologia, 38:693-700, 2000.
- [44] R. K. Yin. Looking at upside-down faces. Journal of Experimental Psychology, 81(1):141-145, 1969.



# Stats & Nets versus Rules & Symbols: Re-opening the Debate on Learning Mechanisms for Articial Grammar Learning

## Marieke S. Woensdregt

Supervisor: dr. Jelle Zuidema Institute for Logic, Language and Computation, University of Amsterdam

#### ABSTRACT

In 1999 Marcus et al. claimed, based on a series of articial grammar learning experiments, that infants must have both a statistical learning and an algebraic rule learning system at their disposal. They were led to this conclusion by the finding that a simple recurrent network model (as an example of a purely statistical learning system) was not able to simulate the generalization abilities of infants. This conclusion sparked a heated debate amongst connectionist modellers on how to overcome this putative shortcoming. As I will argue in this paper however, many of the contributions to this debate only worked to obscure the fundamental question that Marcus et al. based their claim on. This paper discusses this debate both in the light of this fundamental question raised by Marcus et al. and in the light of recent developments in the field of connectionist modelling. Following from this discussion, a new model is presented that combines the simple recurrent network architecture with the concept of encapsulated representations, which is then tested on the Marcus et al. task. The results of these simulations lead us to the conclusion that although the fundamental limitation of neural networks identied by Marcus has to be acknowledged, the claim by Marcus et al. with respect to articial grammar learning may have been too strong.

#### **KEYWORDS**

Artificial Grammar Learning, Connectionism, Statistical Learning

#### **1. INTRODUCTION**

Almost 15 years ago, Marcus et al. (1999) ran a series of experiments to investigate what kind of learning mechanisms infants use to pick up regularities in language. To do so, Marcus et al. tested whether 7-month-old infants could discriminate between two types of artificial grammar sentences. Two types that could be distinguished only on the basis of higher-order dependencies in their structure; not by simple counting or keeping track of transitional probabilities. The results showed not only that infants of this age are perfectly able to make this distinction, but also that they are able to generalize this distinction freely to novel items.

The sentences that Marcus et al. used consisted of three CV words which could occur in different orders. In experiment I and II sentences of ABA order (e.g. ga ti ga) were pitted against sentences of ABB order (ga ti ti ), and in experiment III ABB was pitted against AAB. Each of these experiments consisted of an initial habituation phase in which the infants were exposed to only one of the two sentence orders for two minutes, followed by a test phase in which they were presented with sentences of both orders, but all made up of novel words. Whether infants were able to discriminate between the two orders was determined on the basis of looking times. (In general longer looking times are taken to be indicative of surprisal (Aslin, 2007), but regardless of the direction of the effect, a significant difference between the two sentence orders indicates a difference in processing.) The results showed that during the test phase all but one infant looked significantly longer to the side of the sentences with the novel structure. This led Marcus et al. to conclude that the infants discriminated between the two sentence types, even though both orders were presented with entirely novel words in the test phase.

The fact that these infants could generalize the distinction to sentences with novel items shows that the representation they made of the habituation sentences can not have been bound directly to the words. Rather, they must have represented the input sentences on a more abstract level to be able to recognize that the sentences with the 'old' structure were still the same in the test phase. This ending led Marcus et al. to claim that apart from a statistical learning mechanism, human infants must also have an abstract rule learning mechanism at their disposal. A learning system that describes abstract algebra-like rules (or operations) over symbolic variables. In the case of the present experiment, this would mean that the words would have to be replaced by abstract place-holders, such as 'XYZ', and the relationship between them would have to be described by an algebra-like rule, such as `the first item X is the same as the last item Z' (to describe an ABA sentence). According to Marcus et al., this is the only way to explain the fact that the infants could generalize the structure-distinction to sentences with novel words.

To further support this claim, Marcus et al. ran a series of simulations with a simple recurrent network model (SRN; (Elman, 1990)). Because this is a connectionist model, it works only through statistical learning and not with abstract algebra-like rules. Therefore this model served as the perfect test case to see whether the performance of the infants could be replicated without any algebralike learning. Marcus et al. designed two versions of the SRN (one localist and one distributed) and presented them with the same input sentences as the infants had received<sup>1</sup>. They trained these SRNs on one of the two sentence types, with the task to predict the next word in the sequence. The second word of these sentences cannot be predicted based on the first, but the third word can be predicted based on the first and the second when the structure has been learned (ABA or ABB). The results showed that both types of SRN were able to predict the third word correctly after a reasonable amount of training. However, when the SRNs were tested on the same sentence structure with novel words, their performance dropped to chance level again. Thus, they were not able to perform the generalization to novel items that the infants made. For Marcus et al., this result provided the ultimate evidence that children must possess another type of learning mechanism apart from statistical learning; one that can represent abstract rules independent of the concrete form of the input.

Many connectionist modellers disagreed with this bold conclusion, which Marcus et al. based on simulations with just two relatively simple models. These modellers came up with several different adjustments to network architecture or learning procedure that, as they claimed, solved the problem (Altmann et al., 1999; Seidenberg et al., 1999; Negishi, 1999; Christiansen and Curtin, 1999). However, the conclusion of Marcus et al. (1999) was based on a more fundamental theoretical argument about connectionist models in general, which Marcus advanced in his 1998 paper. The different solutions that were proposed by the connectionistprotagonists did not focus on this more fundamental problem and the debate never came to a conclusive ending.

In this paper I will readdress this problem following the line of argument that Marcus advanced in his 1998 paper, while also staying close to the original question: whether an algebraic rulelearning system is really necessary to distinguish between ABA and ABB structures. I will start by giving a theoretical discussion of the limitation of connectionist models that Marcus uncovered in his 1998 paper, and the implications that this has for the principles of connectionism (section 2). Subsequently I will briefly discuss the different connectionist models that have been proposed to solve the Marcus et al. (1999) task and why they are not fit to answer the fundamental question (section 3). This will be followed by the presentation of a new connectionist model that does address this fundamental question, and the results that were obtained with this model (section 4). This will lead into a discussion that will have a more or less reconciling conclusion; although Marcus (1998) was right in uncovering a fundamental limitation of neural network models. Marcus et al.'s (1999) conclusion with respect to the articial grammar experiment was too strong (section 5).

#### 2. A FUNDAMENTAL LIMITATION OF CONNECTION-IST MODELS?

Marcus et al. explained the fact that their SRN models were not able to generalize to novel words (as the infants were) not as a failure of their particular model, but as an example of a fundamental limitation of neural network models in general. That limitation being that neural network models are fundamentally unable to generalize to stimulus types that they have not been trained on. For example, if the input is represented in the form of individual words, the model will not be able to generalize to words that were not in the training set, if the input is represented in the form of phonetic features the model will not be able to generalize to novel phonetic features, and so on, ad innitum. In his 1998 paper, Marcus et al. coined this problem the problem of generalizing outside the training space. According to Marcus et al., this is a problem that connectionist models can by denition not solve. Below I will review his argument step by step.

#### 2.1 Systematicity

The problem of generalizing outside the training space as put forward by 1998, is part of a broader debate about systematicity in neural networks that was ignited by Fodor and Pylyshyn in 1988. Fodor and Pylyshyn claim that although connectionist models are capable of learning a vast set of regularities, they are unable to make systematic inferences from these regularities to structurally similar but novel items (which is something the human mind does across the board). The most complete description of what they mean by systematicity is given by Fodor and Pylyshyn as "[the phenomenon that] the ability to produce/understand some sentences/thoughts is intrinsically connected to the ability to produce/understand certain others" (Fodor and Pylyshyn, 1988, p. 25).

To illustrate, Fodor and Pylyshyn give the example that when someone knows the meaning of the sentence `John loves the girl',

<sup>1</sup> For the tlearn files of these simulations, visit http://www.psych.nyu.edu/ gary/science/es.html.

#### WOENSDREGT

this person would automatically also know the meaning of `The girl loves John', even if she has never encountered the second sentence before. In other words, the second sentence is treated symmetrically to the first one. And this is not just the case for language. As Fodor and Pylyshyn go on to argue, you will neither find anyone who can think the thought 'John loves the girl', but cannot entertain a mental representation of 'The girl loves John'. The way in which the human cognitive system treats semantically related sentences or thoughts is thus systematic. According to Fodor and Pylyshyn, the only way to account for this systematicity is through making two assumptions: (i) that mental representations have internal structure (consisting of atomic parts that can be combined in certain ways to make up larger wholes), and (ii) that the processes that operate over these mental representations are sensitive to this structure (pp. 12-13)<sup>2</sup>. As Fodor and Pylyshyn argue, these two requirements can by definition not be met by neural networks, since these models work purely through associations between representations. That is, all relationships between representations have to be instantiated by the way individual nodes excite and inhibit each other. This according to Fodor and Pylyshyn has the consequence that neural networks can only instantiate 'causal relationships' and not 'structural relationships' (of which they name constituency as an example).

Where exactly the impediment lies according to Fodor and Pylyshyn remains a bit unclear, but Borensztajn et al. provide a modern analysis of the systematicity debate. They come to the conclusion that in order to account for systematicity, one specific feature is crucial: the existence of substitution classes. These are "classes whose members can all be treated alike", and of which the membership is insensitive to context. According to Borensztajn et al. however, this characteristic can not only be achieved by symbolic models working with rules and variables, but also by neural networks, as long as they allow for encapsulated representations. I will elaborate more on this idea in section 2.5. Let us however first return to the specic systematicity problem that Marcus (1998) identified: the problem of generalizing outside the training space.

#### 2.2 Generalizing outside the training space

To show how humans and neural networks respond differently when confronted with novel items, Marcus presents the example in table 1. Looking at the data in table 1, and given the novel input sequence [11111], any human subject would presumably make the generalization `output is same as input' and would thus output [11111]. However, a neural network would generalize differently. Because the network has never been trained to output 1 on the fifth bit (which would correspond to the fifth node in the output layer), its response to the novel input will be [11110]<sup>3</sup>. This is a result of the fact that the input feature `1 on the fifth bit' never occurs in the training examples and thus lies outside the `training space'.

The boundaries of the training space are dened by the individual features that are present in the training set. A novel stimulus that

was not part of the training set as such can still be within the training space as long as it is made up of features that have all occurred in the training set. If, for instance, the training set consists of [11] and [00], the novel items [10] and [01] fall within the training space, because they can be composed of features that were present in the training set. However, in the example of table 1, the feature `1 on the fifth bit' is not part of any of the training examples, and thus lies outside the training space. What lies inside and outside of the training space therefore depends very closely on how the input is represented. If every node of the input layer represents a full word, all novel words will lie outside the training space. However, if every node in the input layer represents a single phoneme, only novel words that contain novel phonemes lie outside the training space. Similarly, if every node in the input layer represents single phonetic features, only novel words that contain novel phonetic features lie outside the training space, etc. Formally, the training space is thus a subset of the input space that is delimited by the features that have occurred in the training set.

Now, turning back to the output that the network gave for the novel item [11111], it is important to understand that this output is not nonsensical. From the point of view of the network it is a perfectly reasonable induction, for instance because on the basis of this training set, the conditional probability of the fifth bit turning into `1' is 0.0. The important conclusion to draw here however is that the induction that the network makes differs sharply from the induction that a human being would make. Below I will describe the computational cause of this difference.

#### 2.3 Input and output independence

Neural network models learn by adjusting their connection weights on the basis of the training examples that they get as input. In the case of supervised learning, this happens through calculating an error term based on the difference between the output that the network gives and some desired `target' output. This error term is then fed back into the network to determine the direction and degree by which connection weights need to be adjusted, a process known as `error backpropagation'. (See appendix A for a more elaborate explanation.) An important thing to note here is that the error term that is used for backpropagation is always weighted by the activation of the 'sending' node (see equation 2). This means that when the activation level of a certain input node is 0.0, nothing will change in any of the connections originating from this node. Thus, whether the connections originating from a certain input node are changed or not depends on the activation level of that particular input node, but not on the activation levels of the nodes surrounding it. This is what Marcus (1998) calls input independence.

Crucially, the same holds for the training of the output nodes. The equations that determine the weight changes of the connections feeding into a certain output node do not make reference to any aspect of the other output nodes (neither their activation levels, nor their targets, nor the weights of the connections feeding into them, see equations 6 and 10 to verify). Thus, from the network's perspective, every output node is a separate learning problem. This

<sup>2</sup> Note that these are exactly the principles through which the classical symbolic models of cognition work.

<sup>3</sup> This prediction was tested and confirmed by Marcus by running simulations with a simple feed-forward network with one hidden layer.

68

Table 1. A sample function, adopted from Marcus (1998)

Training cases			
Input	Output		
00010	00010		
00100	00100		
00110	00110		
01000	01000		
01010	01010		
01100	01100		
01110	01110		
10000	10000		
10010	10010		
10100	10100		
10110	10110		
11000	11000		
11010	11010		
11100	11100		
11110	11110		

is what Marcus calls `output independence'. The consequence of this input and output independence is that input and output nodes that represent features that were not present in the training set do not participate in the training at all, and the generalizations that the trained part of the output layer has learned to make are not transferred to the untrained output nodes.

Knowing this, it is not surprising that the network trained on table 1 responds with [11110] to the novel input item [11111]. After all, the fifth output node has no idea about the generalization that the other output nodes are implementing. But can we not overcome that problem by simple adjustments in architecture, representation scheme or training procedure?

#### 2.4 Solutions that do not solve the problem

First of all, the introduction of extra hidden nodes or extra hidden layers would not solve this problem. Although extra hidden nodes allow the network to represent more combinations of features internally, and extra hidden layers allow it to exploit more combinations of these combinations of features, the problem of output independence would remain the same. To learn a general function over all output nodes, every single one of them needs to be trained on this function.

Secondly, changing the representation scheme over the input nodes can seem to solve the problem, but this solution is illusory. What a more distributed representation scheme can do is to make more input and output nodes to participate in training. For example, if every input node of a network represents a whole word, a network trained on `ga' and `ti' will have no idea what to do with `ta' and 'gi'. If the input nodes represent phonemes however, 'ta' and 'gi' will not contain any novel features and the network will be able to generalize correctly. This is not really a solution however; it only pushes the problem into smaller corners of the input space. The network would for instance still have no idea what to do with `ba' and 'gu' (because these contain novel phonemes). No matter how we choose the representation scheme, as long as we use a finite input layer there will always be features that lie outside the training space. (Unless the training space covers all of the input space, but then the network would not be facing a generalization task.)

#### STATS & NETS VS RULES & SYMBOLS

A third solution that has been proposed is to give the network more experience. The fact that humans and even infants are such incredible generalization machines might be because of the tremendous amount of experience they gather during life, compared to which the training input of a neural network is often very limited, both in amount and variety. So maybe if the network is first trained on a very simple function (e.g., simply repeating the input) with all items in the input space, so that all input and output nodes will have experience with all input items, and it is subsequently trained on the actual task, then it will be able to generalize correctly. However, Marcus presents two reasons to believe that this is not a real solution either. Firstly, the problem of training independence will not be solved by seeing more training examples of the same kind (because they will not expand the training space in any way). Secondly, if the network is provided experience with items that are not part of the training set for the actual task, Marcus argues that this will not solve the problem either. He comes to this conclusion on the basis of a series of simulations with the network of 1. He first gave this network a pretraining phase in which it was trained on the `identity function' (i.e. simply repeating the input) for all features in the input space until it had learned this function perfectly. Subsequently he trained this pre-trained network on the `reversal function' (i.e. turning [11001] into [10011]), with again leaving one bit out of the training space. Despite of its additional experience, the network still did not learn to generalize this function to input items that involved the bit that was not trained on the reversal function. (And Marcus found the same results with a simple recurrent network trained on a slightly different task.) Marcus ascribes this finding to the fact that even though the `left out' bit lay within the training space of the identity function, it still lay outside the training space of the reversal function. Thus, it is the training space with respect to a certain function that matters.

In sum, Marcus argues that there is one type of generalization that neural networks are by denition not able to make. He coins these generalizations `universally guantied one-to-one mappings'. The term 'universally guantied' refers to relations that hold for all elements in a certain category, and `one-to-one mappings' refer to the situation where every input has a unique output. To relate this back to the task that this paper focuses on, the experiment of Marcus et al., this means the following. If a network has to learn a universally quantied ABB structure and generalize this to novel words that have their own unique output nodes, it does not have any use of knowing what to predict for the words that it was trained on, because the output nodes for the novel words will not have participated in this training. Because the network learns the prediction function (i.e. the sentence structure) for every word independently, prior experience with other words will not help the network to generalize the structure to the new input-output pair. According to Marcus, the real problem thus lies in the backpropagation algorithm itself, not in the particularities of the architecture or the learning procedure.

## 2.5 The principles of connectionism and encapsulated representations

Having identied the limitation described above, it is insightful to

#### WOENSDREGT

review the foundations of connectionism itself, and see how these foundations contribute to this problem. As described in the previous section, "The localism that underlies backpropagation is incompatible with generalizing universally quantied one-to-one mappings to novel items." (Marcus, 1998, p. 268). However, as Borensztajn et al. (2014) argue, this commitment of local associationist processing should not be the core commitment of connectionism. Instead, they argue for a different core commitment to be leading:

"[The interpretations that are permitted of the primitive units of the system] must depend exclusively on internal, autonomously executed processes, and not imposed externally or globally. In other words, all interactions must be local, and all meanings must eventually be grounded in the external input to the network. This constraint rules out symbolic rules with variables, but does not prohibit operations over encapsulated representations (representations not directly affected by processing elsewhere in the system)." (Borensztajn et al., 2014, p. 166)

The encapsulated representations that Borensztajn et al. advocate here open up an interesting new avenue for connectionist modelling. What they envisage is a type of internal representation that is not directly affected by processing elsewhere in the network. In the case of the Marcus et al. experiment, this could mean that the network holds an internal representation of the word category A that responds the same regardless of whether it binds to word A1, A2 or A3. This would provide a means to represent input structures (such as A-B-A) 'explicitly' in the network, which according to Fodor and Pylyshyn (1988) is the fundamental requirement for systematicity (see section 2.1). These encapsulated representations would be similar to variables in the sense that they can act as placeholders that can bind to different input units. However, they are still different from variables in the sense that the scope of input units that they bind to will still have to be learned through experience, rather than being specied globally in advance (as would be the case in a classical symbolic model). The crucial innovation that these encapsulated representations would bring into neural network modelling is that they would allow the categories that emerge inside the network to play a causal role in its behaviour, whereas in earlier (eliminative) connectionist models the categories that emerged were only appreciated by an external observer (i.e. the modeller) and not `visible' or usable to the network itself.

In sum, what Fodor and Pylyshyn argued for is that in order to perform systematic inferences a model needs to have explicit categories, leading to representations of the input structure that the model can act on. According to Fodor and Pylyshyn this can only be achieved in a symbolic model that works with variables, but according to Borensztajn et al. this is also possible in a connectionist model by means of encapsulated representations.

Would this new notion of encapsulated representations be able to solve the problem of generalizing outside the training space? Not necessarily, because the scope of the category (i.e. which input and output nodes it binds to) would still have to be learned. However, the existence of encapsulated internal categories might overcome the problem of generalizing from one function to another. To clarify this intuition, say we would first train a neural network on the identity function to form an encapsulated internal category 'B' that contains slots for five bits. We could then train it on the reversal function with a certain training space, and subsequently train it on the identity function again with some novel five-bit sequences that lay outside the training space of the reversal function. The network would now have learned that all inputs it has seen are part of the category B, and it has learned how to reverse whatever binds to category B. Together, this should allow the network to generalize the reversal function correctly to novel items, even though they lie outside the training space of the reversal function. In this way, Marcus's statement that it is the training space with respect to a particular function that matters, could then be overcome.

#### 3. THE PROPOSED CONNECTIONIST SOLUTIONS

Before exploring the possible solution provided by encapsulated representations however, I will first review the solutions that other connectionist modellers proposed in response to the publication of Marcus et al. (1999).

#### 3.1 The original simulations (1999)

As described in section 1, Marcus et al. (1999) kicked off the debate by attempting to model the results they obtained with 7-month-old infants with two different versions of the SRN. The first of these two versions was a localist network that contained one input node for every word that occurred in the training set. This yielded an input layer of 13 units: one for each word (12 in total) and one punctuation marker. The second was a distributed network that represented each word as a combination of six phonetic features (using the distinctive feature notation of Plunkett and Marchman (1993)). This encoding scheme yielded an input layer of 13 units: two phonemes for each word, six features for each phoneme and one punctuation marker. In both versions of Marcus et al. (1999)'s model, this input layer was connected to a hidden layer of 40 nodes (with a corresponding context layer of 40 nodes). This hidden layer was connected to an output layer of 13 nodes that corresponded to the encoding scheme used for the input layer.

Marcus et al. trained both versions of their network on a prediction task; for every input word, the network had to output the next word in the sequence (where all input sentence had the ABB structure). Only the third word in such sentences can be predicted based on the learned structure and the rst and second word, and indeed both networks were found to be able to predict this third word correctly after a reasonable amount of training. However, in a subsequent test phase, both networks were found unable to generalize this learned structure to novel words. To give a concrete example, given `leje..' (part of the training space), the localist network was able to predicted `li' (part of the training space) with an activation level of < 0.01. Thus, both the localist and distributed version of this SRN were found unable to generalize outside the training space (as expected, given the

#### analysis of Marcus (1998)).

# 3.2 Adding experience and changing the task to categorization: Seidenberg & Elman (1999)

The contribution of Seidenberg and Elman to the collection of commentaries in (Seidenberg et al., 1999) describes how they took Marcus et al.'s distributed SRN and made two essential adjustments to the training and testing procedure. Firstly, they granted their model an extensive pretraining phase for it to gain experience with all possible words (i.e. the whole input space). Secondly, they changed the task on which the network was trained from prediction to categorization.

In their pretraining phase, Seidenberg and Elman exposed their network to a set of 120 CV-words, chosen such that they contained the full range of phonetic contrasts of the English language. This means that the pretraining stimuli covered the entire input space of this experiment. For every word the network was trained on what we could call a `similarity' function. It was trained to output whether the current word was the same or different compared to the previous one. After the network was exposed to 6\*50,000(= 300,000) of these words, the setting of connection weights was saved and the network entered the categorization training phase.

To change the task from a prediction into a categorization task, Seidenberg and Elman presented the network with both sentence structures (i.e. ABA and ABB) straight away, and trained it to distinguish these two types of sentences. This output was given on a different output node than the one that was used in the pretraining phase, to allow the connection weights between the hidden layer and the output node to be trained independently of the function that was learned in pretraining. After having seen the total set of 32 sentences for 347 times (=11104 trials in total), the network had learned to discriminate correctly between the two types of sentences.

This training phase was followed by a test phase in which the network was presented with four sentences that were made up of four entirely novel words (neither part of the pretraining nor of the categorization training phase). Crucially, the phonetic features that these novel words consisted of had all been covered during pretraining, but not during categorization training. That means that the test items lay within the training space of the similarity function, but outside the training space of the categorization function. Nonetheless, the results showed that the network could generalize its categorization function correctly to these novel items.

So maybe generalization from one function to another, even when the training space of the second function is only a subset of the training space of the first function, is possible after all? To answer this question, we have to look more closely at the network that Seidenberg and Elman used, which is depicted in Figure 1.

As we can see in Figure 1, the model that Seidenberg and Elman used has only two output nodes. The first nodes is trained during pretraining to answer `1' if the current word is identical, and `0' if it is different. The second node is trained during the categorization training and testing phases to answer `1' if the sentence has an

#### STATS & NETS VS RULES & SYMBOLS

ABB structure and `0' if the sentence has an ABA structure. During the pretraining phase, the network learns to represent some kind of meaningful combination of the representations of the current and the previous word in the hidden layer, which then allows the output node to answer `1' if they are the same and `0' if they are different. To adjust this similarity judgement network into a categorization network, all that needs to be done is to output `1' if the second and third word are the same (= ABB structure), and `0' if they are different (= ABA structure). Input independence is not a problem any more at this stage because the network has already learned to make a meaningful mapping from all input units to the hidden layer during the pretraining phase. And crucially, output independence is not a problem here either because the network only has to learn a mapping from the hidden layer to one output node, so it does not have to learn any mappings to new output nodes in response to the novel input. One can expect that if this network would have to give its output on novel nodes in the test phase, it would not be able to make this generalization. Therefore, the results of this model do not provide any evidence against Marcus (1998)'s claim that neural networks are not able to generalize universally quantied one-to-one mappings.

Another important point of critique to this model was pointed out by Marcus in his commentary to the contribution of Seidenberg and Elman in (Seidenberg et al., 1999)s. As Marcus argues, Seidenberg and Elman have abandoned one of the principles of connectionism (that we identied in section 2.5) by changing the task from prediction to categorization. In Seidenberg and Elman's pretraining procedure, the data for the similarity function is not directly observable from the environment. To clarify this statement, compare it to the prediction task that the networks of Marcus et al. (1999) were trained on. Here, determining whether the output is correct or incorrect can be done simply by looking at what the next input word is and whether the output activation corresponds to this or not. Thus, the data for the function that the network is trained on is directly observable from the environment (i.e. the input to the network). In the case of Seidenberg and Elman's similarity training however, determining whether the output on the `same/different node' is correct cannot be done on the basis of the direct input to the network. (If the task is to determine whether 'ba' and 'di' are the same, having 'di' as the next input word will not provide the answer.) That means that Seidenberg and Elman have implicitly built in an 'external teacher' that decides whether two words are the same or not, to provide the network with its feedback. To be able to provide this answer for every input sentence, this external teacher has to implement a rule that operates over variables, thereby abandoning the core commitments of connectionism. (See for the continuation of this discussion Elman and Seidenberg (1999) and Marcus (1999b).)

#### 3.3 Extending the input representations: Negishi (1999)

A different type of solution to the problem of simulating the Marcus et al. (1999) results was provided by Negishi (included in the collection of commentaries in (Seidenberg et al., 1999)) and Christiansen and Curtin (1999). Both their models preserved the standard architecture

#### WOENSDREGT

and training procedure of the SRN, but were provided with different input representations.

Negishi took the phonetic feature representations that Marcus et al. used in their distributed version of the model, but turned these features into continuous rather than discrete values. His results showed that during testing, the model made larger prediction errors for the `consistent' sentences (the structure that the model had been trained on) than for the inconsistent sentences, which Negishi took as evidence that the model could discriminate between the two sentence structures. However, as Marcus pointed out in his commentary in the collection (Seidenberg et al., 1999), turning the input nodes into units that can represent features of the input on a continuous scale turns these nodes into variables, i.e. placeholders that can bind to different elements in the environment. Therefore, although interesting, this is again a solution that implicitly implements symbol-manipulation.

Christiansen and Curtin also argued that there is more to speech than just discrete phonetic features, and that the statistical knowledge that the 7-month-old infants of the Marcus et al. study had already acquired in order to segment speech might have helped them make their generalizations. To grant this kind of statistical knowledge to the model as well Christiansen and Curtin used an already existing model of early infant speech segmentation (Christiansen et al., 1998) and trained it on the Marcus et al. data. The model that they used was similar to the distributed SRN of Marcus et al., except for two aspects. Firstly, the input and output layer contained two extra units to represent the stress pattern of the word (no stress, secondary stress or primary stress). Secondly, the nodes of the output layer (but not the input layer) represented complete phonemes rather than phonetic features. Thus the representations in the input and output layer had a different level of detail or grain. Christiansen and Curtin's network was then trained on predicting the next phoneme in the sequence, rather than the next word.

Christiansen and Curtin's results showed that the model was better at segmenting the inconsistent test items, which they interpreted as showing that the inconsistent test items were more salient for the network, analogical to the nding that the infants of Marcus et al. looked longer at the inconsistent items. In this interpretation, the work of Christiansen and Curtin shows that the results obtained with the infants can still be based on statistical learning (of speech segmentation). However, in his reply Marcus (1999a) argues that with their changing of the input representation, Christiansen and Curtin changed the input into something that diered fundamentally from the input that Marcus et al.'s infants were exposed to. Instead of the segmented streams of words that the infants received (with 250-ms gaps between each word), Christiansen and Curtin exposed their model to a continuous speech stream. The results that they acquired are thus based on a segmentation task, which is crucially different from the task of Marcus et al..

#### 3.4 Continued learning during test phase (2013)

During the 2013 CogSci conference in Berlin, Willits (2013) presented the results of a series of simulations with a simple localist

SRN, among which were simulations of the Marcus et al. (1999) experiment. As Willits argues, it is not surprising that the models in the Marcus et al. simulations are not able to generalize correctly to the novel items, because the novel items were not present in the training set. This is exactly the same explanation as the one that Marcus (1998) gives for such ndings. However, according to Willits there is no reason to not allow the network to continue adjusting its connection weights during the test phase (an option not discussed by Marcus). This solution does not contradict Marcus's claim that neural networks are not able to generalize outside the training space of a particular function, because it simply expands the training space for the function of interest. However, this testing procedure is still interesting to consider because it may reveal that the network has in fact stored useful information about the sentence structure somewhere internally, which might only become apparent when some learning in the test phase is allowed.

To investigate this hypothesis, Willits trained a localist SRN (with each input node representing a whole word) on either the ABA or ABB structure, and subsequently subjected it to a `test phase' in which it was trained on novel sentences that were either consistent or inconsistent with the training structure (thus the network was allowed to continue adjusting its weights during this test phase). What Willits showed was that during this test phase, the network learned to predict the novel sentences with consistent structure more guickly than the inconsistent novel sentences. This suggests that the network has in fact stored some information about the abstract structure underlying the training sentences, which it can exploit when it is allowed to continue learning in the testing phase. To check whether this was really the case, Willits analysed the weight conguration of the network after training, which showed that the network's recurrent and output connections had learned to make an abstract representation of the structure of the input sentences. (How exactly this was visible in the weight congurations is however not specied in the conference proceedings.) The conclusion of Willits is therefore that all this network has to do in the test phase is to learn a mapping from the new input nodes to the abstract structure that has already been acquired internally, which allows it to adjust to the consistent test sentences guicker than to the inconsistent test sentences.

# 3.5 Extra hidden layer and continued learning during test phase (1999)

Dienes et al. (1999) designed a model originally not for the purpose of simulating the Marcus et al. (1999) task, but for simulating transfer of learning between different domains (for instance between letter sequences and musical sequences). In order to achieve this kind of transfer in a neural network, Dienes et al. took the standard SRN architecture and adjusted it in two ways. Firstly, they gave their network two input layers and two output layers instead of one. The first input and output layer were used to encode the first domain, and the second input and output layer to encode the second domain. Secondly, Dienes et al. added an extra hidden layer (which they called `the encoding layer') between the input and hidden layer. The



Figure 1. The model that Seidenberg and Elman used. Adopter from http://crl.ucsd.edu/~elman/Papers/MVRVsim.shtml

resulting architecture is shown in Figure 2.

In the training phase, this model is presented with sequences of a particular structure in only one domain, and trained to predict the next element in the sequence. This training phase is followed by a 'classication phase' in which the network is presented with both consistent and inconsistent sequences, to verify that the network has learned (if it has, it should be better at predicting the consistent sequences). Crucially, before the network enters the test phase, the connection weights between the encoding layer and the hidden layer (indicated as the `core weights' in gure 2) are frozen. This means that from now on only the weights indicated in gure 2 as 'mapping weights' can be adjusted. Finally, in the test phase, the network is presented with only sequences from the new domain, both consistent and inconsistent, and classication is tested in the same way as in the classication phase. During this test phase, the mapping weights are allowed to continue learning. Thus, the learning procedure of Dienes et al. is similar to that of Willits (2013), with the only dierence that in Dienes et al.'s simulations, only a subset of the connection weights is allowed to continue learning in the test phase.

To simulate the Marcus et al. experiment, Altmann et al. trained the Dienes et al. model first only on one sentence structure (corresponding to the `first domain'), with each input node corresponding to a particular word (i.e. the localist encoding scheme of Marcus et al.). This model was subsequently tested on both types of structures (i.e. both `domains') with novel sentences (using input nodes that had not been used during training). The results showed that the network was now signicantly better at predicting the next word in the consistent sentences than in the inconsistent sentences. Altmann et al. took this as evidence that the network could successfully discriminate between the two sentence structures.

In his response (published together with the technical comment of Altmann et al. (1999)), Marcus gives three main points of critique to this model. The first point has to do with the method of analysis that Altmann et al. (1999) use, which diers from the standard analysis of SRN performance. When analysing these results with the standard method however, Marcus shows that the model has not really learned the grammar that it is trained on, but rather that it simply oscillates between the ABA and ABB grammars. The second point has to do with the cognitive realism of the model. The results that were obtained depend on the `freezing' of connection weights during the test phase, and on iterating over each test item several times in a row. It is unclear how the brain would perform such freezing of weights, and how the infantsin the Marcus experiment would be able to iterate over the test items multiple times, given the very brief test period in the experiment (only 12 sentences). The third and most important point of critique relates to what it is that the network learns in the Altmann et al. task. According to Marcus, the network might just be learning a mapping from one set of words to another, rather than learning an abstract representation of the sentence structure.

To see if this was indeed the case, Marcus conducted new simulations with the Altmann et al. model with a slightly different test phase. Marcus first trained the model on ABA sentences (e.g. 'ga ti ga'), followed by a test phase consisting of an ABA sentence with novel words (e.g. 'wo fe wo'). When the model had learned the correct mapping from `ga ti ga' to `wo fe wo', Marcus subsequently tested it on either 'fe wo wo' or 'fe wo fe'. Using the method of analysis of Altmann et al., it turned out that the model now had more difficulty predicting `fe wo fe' than `fe wo wo'. Presumably this is because the model has learned that 'wo' always appears in the third position, rather than having learned that whatever word comes first in the sentence will also be in the third position (the abstract ABA representation). This is surprising finding if we assume that the model has learned an abstract representation of the sentence structure, but it is not if we assume that the model has just learned a mapping from one set of words to another set of words. However, it is important to note that this precise procedure of training and testing did not occur in the original Marcus et al. experiments, so it remains an open question how human infants would respond in this condition.

Interestingly, the model of Dienes et al. (1999) and Altmann et al. (1999) in a way implements the idea of encapsulated representations discussed in section 2.5. Freezing the core weights


Figure 2. The network architecture used by Dienes et al. Adopted from Dienes et al. (1999).

of figure 2 renders the internal representations that have arisen in the hidden layer during training insensitive to context during the test phase. Whether the input is from the first or the second domain (or inside or outside of the training space), the representations between the encoding and hidden layer remain the same. The network only learns a new mapping from the novel input items to these internal (encapsulated) representations, and from these internal representations to the output units. Thus, if the network would be able to come to an abstract representation of the sentence structure that it is trained on, it might not have too much diculty generalizing this to novel input items. Even if these lie outside the training space of the prediction function.

In the remainder of this paper, I will present a new model that implements exactly this idea, by means of the minimally necessary changes to the models and procedures that Marcus et al. (1999) used. With this new model, I will first explore the question how an abstract representation of the input grammar might be acquired, to then address the question how encapsulation of this representation by weight freezing affects further training and testing.

#### 4. A NEW MODEL: USING ABSTRACT STRUCTURE REPRESENTATIONS IN A CAUSAL WAY

As described in the previous section, this new model was developed to explore a way of getting a neural network model to acquire an abstract representation of an ABA or ABB sentence structure, and to have this representation play a causal role in the behaviour of the network.

#### 4.1 The architecture

The architecture of the model presented here is based on the intuition that for the network to come to an abstract representation of the sentence structure, it should first be able to represent two words at one time step, to then compute something about the relationship between the two (e.g. same/different). In the standard SRN, the

hidden layer is the layer that can contain information about both the current and the previous input item (e.g. word two and word three) at the same time. However, to really exploit this information the network should be able to perform a computation over the combination of these two items (such as subtraction). This computation could be performed by the connections between the hidden and the output layer, but then the output layer would not contain a prediction that can be interpreted by the outside world (i.e. a representation that corresponds to the way the input is represented). Therefore, in the current model an extra hidden layer is inserted between the `recurrent' hidden layer and the output layer. Following Dienes et al. (1999), I call this layer the `encoding layer'. The resulting network architecture is shown in Figure 3.



Figure 3. The architecture of my model: a standard SRN with an extra hidden layer between the normal hidden layer and the output layer. The connections encircled in blue are the encoding weights that are frozen in section 4.5.

To achieve the effect of encapsulated representations as described in section 3.5, the connection weights that are encircled in blue in figure 3 are the weights that will be frozen after training (see further section 4.5).

## 4.2 Results I: Replication of the Marcus et al. (1999) simulations

Firstly, I replicated the results of Marcus et al. (1999) using my own implementations of the two versions of their model. These results are thus not obtained with the network depicted in gure 3, but simply with the same architecture, same data, and same training and testing procedure as Marcus et al. used, as depicted in figure 4.



Figure 4. A general impression of the network used by Marcus et al. (1999)

To measure the network's prediction I used a `nearest neighbour' measure, simply taking the word of which the difference between

#### STATS & NETS VS RULES & SYMBOLS

its vector representation and the output vector was smallest as the 'predicted word'. (For example, is 'ba' is represented as [1000] and 'po' is represented as [0100] the output vector [.7.2.4.3] is analysed as predicting the word 'ba'.) Using this prediction measure, the output of the network was then scored as either correct or incorrect, and the average number of correct predictions out of a set of 16 test items over 4 runs of the simulation was used to yield an average percentage correct score. In addition to this average percentage correct score, an average error score was also computed based on the error term that the network computes internally as part of the backpropagation algorithm. This is a root mean squared error score, as depicted in equation 1. This RMS error measure was recorded at every third word of each sentence, which allows us to track the improvement that the network makes during training.

$$\sqrt{\frac{\sum (target - output)^2]}{n}} \tag{1}$$

My implementation of the localist version of Marcus et al.'s network (with a learning rate of 0.5 and a bias weight of 0.0), reached an average percentage correct score of 37.50% after being trained for 230 epochs on a training set of 48 ABB sentences (= 11040 trials in total). Chance level in this case is in principle 7.69% (i.e. 1 out of 13), because there are 13 possible outputs. However, because the network has only seen 9 out of 13 possible input words at this stage, a more conservative chance level (and more realistic, considering output independence (Marcus, 1998)) would be 11.11% (1 out of 9). In both cases however, the average score of the network is suciently high above chance level to assume that the network is doing something sensible here. It can be expected that with more training the network would be able to learn this function perfectly. However, when we freeze the setting of all connection weights after this training period and then subject the network to a test phase with novel sentences, the performance of the network drops to chance level again. (Following Marcus et al., the test set consisted of two sentences containing four completely novel words (corresponding to four previously unused input nodes). In all runs of the simulation, the model now got `stuck' on predicting one single word that was part of the training set for every single test item, yielding an average score of 0.00% correct on the test items.

The development of the RMS error score, in the training phase compared to the test phase is depicted in figure 5 (with mean and standard deviation averaged over 4 runs). In this graph the training and test phase are depicted alongside each other to allow for a direct comparison, with the training phase in blue and the test phase in red. As figure 5 shows, the RMS error gradually goes down during training whereas in the test phase it shoots up to a xed value even higher than the value at time 0 of the training phase, when the network was initiated with random weights.

For the distributed version of my implementation of this model, the measure of accuracy needs to be slightly different. Because this network has six output nodes (corresponding to six phonetic features) to represent the first phoneme of the word it predicts, and six different output nodes to represent the second phoneme it predicts, this network can make partial errors in its output. It can for instance output `le' when the desired output is `li'. When this was the case this was scored as a `0.5 correct' answer, reflecting the interpretation that the network has learned to predict one of the two phonemes correctly. After being trained on the ABB structure for 500 epochs of 48 sentences (= 24000 trials in total), again with learning rate = 0.5 and bias = 0.0, this network reached an average percentage correct of 40.63%. Again at this stage the conservative chance level is 11.11%, and the average value of 40.63% is suciently high above this level to assume that the network has learned something sensible. The RMS error graph resulting from these simulations looked similar to the one depicted in figure 5.



Figure 5. Development of RMS error score over training phase (blue) compared to test phase (red) in replication of Marcus et al. (1999)'s localist network. Mean and standard deviation over 4 runs.

Thus, for this version the same conclusion holds as for the localist version; its performance is sufficiently high above chance level to assume that with longer training it could learn the prediction function perfectly. However, if we then subject this network to a test phase with two entirely novel sentences that both contain one feature that lies outside the training space (again following exactly the procedure of Marcus et al. (1999)), performance drops to chance level again (an average of 0.00%).

In sum, my implementation of the Marcus et al. simulations replicates exactly the lack of generalization ability that was found by Marcus et al.. Both the localist and the distributed version of this network can learn to predict the third word in the sequence after training on one particular sentence structure, but when tested on sentences that contain features that lay outside the training space, they are both completely unable to generalize. Importantly, these results are exactly as is to be expected given the analysis of Marcus (1998) as described in section 2. Because the output nodes that represent the novel input features that are introduced in the test phase have not participated in training, the prediction function that the network has learned cannot be generalized correctly to these novel features. The network has no clue how to relate the internal representation that it may have acquired to its as yet untouched input units and corresponding untrained output nodes.

#### WOENSDREGT

# 4.3 Results II: The effect of an extra encoding layer for representation potential

Now let us turn to the results obtained with the new model presented here. The first question that should be asked when examining a new model architecture is whether it is able to represent the solution that the designer hopes it to nd. This is a step that is often skipped (or at least not reported) in the research of connectionist modellers, leading to the kind of simulation results as the ones reported by Marcus et al. (1999). These training results could have simply been predicted based on the characteristics of the training and testing procedure (following Marcus (1998)), and they obscured an internal representational potential of the model, as was revealed by Willits (2013). Thus, before looking at what the model can acquire through training, it is interesting to look at what the model can achieve when implemented with a handcrafted `ideal' setting of the connection weights. In gure 6 such a handcrafted version of the current model is depicted, with an ideal setting of weights for predicting the ABA structure. If we turn this general depiction of the network into a localist version to solve the Marcus et al. task, it would contain 13 nodes in the input layer (corresponding to individual words), 26 nodes in the hidden layer, 26 nodes in the context layer, 13 nodes in the encoding layer and 13 nodes in the output layer.



Figure 6. A handcrafted solution for the ABA structure.

Looking at the labels 't', 't-1' etc. in figure 6, the reader can follow the flow of information from input to output layer. In this handcrafted version, the activation levels of all 13 input nodes are first copied one-to-one to the right-hand side of the hidden layer. All activation levels of the hidden layer are then copied one-to-one to the context layer. At the following time step, the right-hand side of the context layer (which corresponds to the right-hand side of the hidden layer at the previous time step) is then copied one-to-one to the left-hand side of the hidden layer. Consequently, the hidden layer now contains the full input representation of the word at time t-1 in its left-hand side, and the full input representation of the word at time t in its right-hand side. The word at time t-1 is then copied to the 13 nodes of the encoding layer with one-to-one mappings of weight -1, whereas the word at time t is copied to the same 13 nodes of the encoding layer with one-to-one mappings of weight +1. This has the effect that the encoding layer now holds a subtraction of the word at time t from the word at time t-1, which leaves the representation of the word at time t-1 to remain in this layer. Note that because all copying was done with exact one-to-one mappings, the encoding layer now contains a representation corresponding in form and detail to the input representations (consisting of 13 nodes that correspond to 13 different words). These activations then only have to be copied one-to-one to the output layer, yielding the word at time t-1 as the answer. In sum, given `le' at time t-1, and `je' at time t, the network will predict that the next word will be `le'.

One thing to note here is that such one-to-one copying of activation levels would in a real simulation be distorted by the sigmoid activation function that modies the flow of activation between layers (see appendix A for more information). This sigmoid activation function (see equation 3) has as its effect that it pushes the activation levels of nodes more towards the two extremes (in the current implementations 0.0 and 1.0), to keep the distinctions between levels of activation clear in all layers. The `sensitivity' of the sigmoid activation function (i.e. around which value of x it centres) can be modied using the weight of the bias node, as is depicted in gure 12. Therefore, whether the weight conguration described above and in gure 6 really works to solve the task depends on the setting of the connections from the bias node. A bias weight of 0.4 was found to render the desired results.

As predicted, a simulation with the network and weight conguration of figure 6, with the bias weight set at 0.4, showed that this network could predict every third word in all ABA sentences correctly. To change this network into an ABB predictor, the only thing that needs to be done is to switch around the setting of weights between the hidden and encoding layer. The weights connecting the left-hand side of the hidden layer to the encoding layer need to be changed from -1 to +1, and the weights connecting the right-hand side of the hidden layer to the encoding layer need to be changed from +1 to -1. This network also showed perfect performance with a bias weight of 0.4, on all ABB sentences. Thus, we can conclude that the new network architecture presented here can at least represent a solution to the problem, where the network performs the same computation internally irrespective of what it gets as input, and thus reproduces the input representations faithfully regardless of their form. The second and most interesting question however is whether this network can also learn such a solution through training with the error backpropagation algorithm.

# 4.4 Results III: The effect of an extra encoding layer for learning a solution

To see whether the model presented here can also learn a solution to the problem through error backpropagation, this model was first subjected to the same training and testing procedure that Marcus et al. tested their network on. That is, only allowing it to learn during the training phase and not during the test phase.

Under these conditions, the localist version trained on ABB sentences reached perfect performance after an average of 5265 training sentences (again with learning rate = 0.5 and bias = 0.0). However, when confronted with the test sentences, its performance dropped to chance level again (0.0% correct). The development of the RMS error levels in the training compared to test phase of these simulations are depicted in figure 7.



Figure 7. Development of RMS error score over training phase (blue) compared to test phase (red) when all weights are frozen before testing. Mean and standard deviation over 4 runs.

Similar results were found when training and testing the model on ABA sentences, although here the model needed a bit more training to reach perfect performance on the training items. Together, these results show that what the network learns with this training and testing procedure is not the generalist solution envisaged in the handcrafted version of figure 6, but rather a more context-sensitive ('input-specific') solution that suffers from the problem of not being able to generalize outside the training space.

# 4.5 Results IV: The effect of continued learning with frozen encoding weights

Although we have seen that this new model is not able to generalize the prediction function that it learns any better than the Marcus et al. model, it is nevertheless interesting to see whether the network has acquired an internal representation of the structure that is at least to some extend abstract and `context-insensitive'. This can be revealed by allowing the network to continue adjusting its connection weights during the test phase, as demonstrated by Willits (see section 3.4). To make sure however that the part of the network that we would expect to hold such an abstract representation, i.e. the connections between the hidden and encoding layer, do not change, these weights were frozen before the start of the test phase (as depicted in figure 6).

The results of evaluating the model with this training and testing procedure (again trained on ABB sentences) is depicted in figure 8. What this gure shows is that the RMS error goes down much quicker in the test phase than it does in the training phase. During

the test phase with the frozen encoding weights the network needs only approximately 1500 sentences to get to the same error level for which it needs more than 8000 trials during training.



Figure 8. Development of RMS error score over training phase (blue) compared to test phase (red) when only the encoding weights are frozen before testing. Mean and standard deviation over 4 runs.

What we can conclude from this is that whatever the internal structure representation is that the network has learned during training, it helps to speed up the subsequent learning in the test phase by more than five times.

#### 4.6 Results V: The effect of the encoding layer during learning

Although the results described in section 4.5 are interesting, similar results were already found by Willits (2013) with a normal SRN without the extra encoding layer (see section 3.4). The question thus arises whether the innovation of adding an extra encoding layer to the network actually helps the network to acquire a more abstract internal representation of the input structure, or whether the results we nd here are simply a replication of the results found by Willits. In order to answer this question, I reran the simulations of Figure 8 but allowed all weights to continue learning during the test phase, and compared this to a replication of Willits's simulation (same procedure but with normal SRN). The results of this comparison are depicted in figure 9.

As figure 9 shows, the addition of the encoding layer actually makes the network slower in improving its performance during the test phase than when no such layer is added. This comparison thus reveals that the extra encoding layer actually slows down the process of nding a new mapping from the novel input to the internal representations to the novel output. This provokes the suspicion that the new model proposed here does actually not acquire a more abstract representation of the input structure by itself, at least not through the training procedure of error backpropagation.

#### 5. DISCUSSION

In this paper I investigate whether a connectionist model could be able to learn an abstract structure representation of such input sequences as the ones used in the experiment of Marcus et al. (1999),



Figure 9. Development of RMS error score over training phase (blue) compared to test phase (red) when none of the weights are frozen before testing, compared to a replication of the Willits (2013) simulation (green). Mean and standard deviation over 4 runs.

and whether this can be represented in such a way that it allows the network to generalize the acquired structure to entirely novel stimuli. This question relates to the debate about a supposedly inevitable limitation of connectionist modelling that Fodor and Pylyshyn (1988) identied as the problem of `systematicity', and Marcus (1998) as the problem of `generalizing universally quantied one-to-one mappings'. This question is very relevant to the debate on what learning mechanisms are minimally required to learn language.

As described in the introduction, Marcus et al. claim that their combined ndings (i) that 7-months-old infants can readily generalize a familiarized ABA or ABB structure to entirely novel items, and (ii) that their SRNs failed to show this generalization ability, demonstrate that human learners must have an `algebraic rule-learning system' at their disposal. This learning system would be present alongside a statistical learning mechanism, and would presumably be innate. However, this claim relies on the assumption that connectionist models are by denition not able to perform such `generalization outside the training space', and that a system based on rules and variables is the only way to solve this task. The latter assumption is what the current paper challenges.

Firstly, what my results show are a replication of the ndings of Marcus et al.. If a regular SRN (without the extra encoding layer) is trained on these data and not granted continued learning in the test phase, it is not able to generalize its prediction function to novel items that lie (partly) outside the training space. This is as to be expected on the basis of the theoretical analysis of Marcus (1998).

The new model that is presented in this paper is an extension of the regular SRN with an extra hidden layer (the `encoding layer') that is located between the recurrent hidden layer and the output layer. As the handcrafted solution described in section 4.4 shows, this extra encoding layer in theory allows the network to form an abstract representation of the input structure that is insensitive to context and can operate as a substitution class, sensu Borensztajn et al. (2014). Thus, this new model is at least able to represent a solution to the problem of generalizing outside the training space. The next question that needs answering however is whether this model can also learn this solution through training with the error backpropagation algorithm. A first evaluation, with exactly the same training and testing procedure as Marcus et al. (1999) used in their simulations, showed that in spite of the extra encoding layer this new model was not able to generalize the function it learned during training to the novel test items.

However, this training and testing procedure, where all connection weights are frozen before the network enters the test phase, cannot give a conclusive answer to the question whether the model can learn an abstract representation of the input structure. Precisely because of the problems with generalizing outside the training space as identied by Marcus (1998), testing the model in this way will not exhaustively reveal all information about the input structure that the network has stored internally. As Willits (2013) argued, there is nothing against allowing the network to continue learning during the test phase, to reveal such potential internal representations.

Willits' method to test this was to compare the speed with which the network learns to predict consistent novel sentences in the test phase to the speed with which it learns to predict inconsistent novel sentences. In the current study I compared the speed of learning the prediction function on the training items to the speed of learning to predict the novel test items (all containing one or more novel input features).

In addition to the architectural innovation of the new model presented here (the extra encoding layer) a procedural innovation was also explored, in the form of the freezing of connection weights. Inspired by the work of Dienes et al. (1999), the intuition behind the freezing of the 'encoding weights' was that it would induce encapsulation of the internal structure representation that the network has learned, in the sense of Borensztajn et al. (2014). This would allow the network to form abstract representations of the word categories 'A' and 'B', which could function as substitution classes, on which the operations within the network could then act. As argued in section 3.5, this could help the network to generalize the function that it has learned during training and make the internal representations less susceptible to changes in the structure of the input. Although I will not claim that the `freezing' of connection weights is a method that the brain also implements, this method does instantiate a simple way of implementing the `encapsulated representations' that Borensztajn et al. (2014) argue for, and which they show are implemented in the brain by means of the columnar organization of the cortex.

The results of this second evaluation, where the `encoding weights' were frozen before testing and the rest of the weights were allowed to continue learning, revealed that the network did learn something valuable about the input structure. After having learned to predict the training items, the network learned to predict the novel test items about ve times as quickly. However, the rst evaluation (where all weights were frozen before testing) showed that the solution that the model had learned during training was not as abstract and generalizable as the one implemented in our handcrafted version (for then it should have been able to generalize to the novel test items without any problem). Thus, the question arises whether this new model really learned to use the encoding layer in the way that was envisaged when designing the model.

To investigate this, a direct comparison was made between this new model and the regular SRN (i.e. without the extra encoding layer). The average speed of learning to predict the novel items after having been trained on the training items was compared between these two models. Although no statistics were computed for this comparison, the resulting graph shows that the average speed up of learning the test items compared to the training items is possibly even greater for the regular SRN. This conclusion would have to be corroborated by statistics, but if it is a real effect this means that rather than helping the network form a more abstract representation of the input structure, the extra encoding layer only slows down the learning process.

Thus, we have to conclude that although the addition of the encoding layer would in theory allow the network to form more abstract representation, which could then be `protected' by means of weight freezing (encapsulation), the model in practice does not nd such a solution, at least not when trained with the error backpropagation algorithm.

An intriguing possibility that remains is that the model presented here would be able to find such a solution if it were initialized with a bias towards one-to-one mappings, because these are an essential aspect of the handcrafted solution presented in section 4.4. Such a bias would also be arguable from a neurobiological point of view, since several parts of the brain work with receptive fields, where a given neuron is only sensitive to a restricted set of other neurons, or a restricted portion of the sensory input. Unfortunately however, this possibility has not been explored here due to lack of time.

Altogether, there are still some caveats in the simulations presented here that would be interesting to fill, but there are some conclusions that we can already safely draw. Although Marcus's (1998) point that neural networks cannot generalize universally quantied one-to-one mappings still stands, the conclusion that Marcus et al. (1999) draw with respect to the learning mechanism that infants must have at their disposal is too bold and cannot be concluded from the simulation results that they present. Without ignoring or obscuring the problem of generalizing outside the training space, as many other responses to the claim of Marcus et al. have done, the current analysis and simulations have shown that the possibility of encapsulating internal representations may potentially enable connectionist models to overcome this limitation. This possibility deserves further exploration.

#### REFERENCES

Altmann, G. T. M., Dienes, Z., and Marcus, G. F. (1999). Rule Learning by Seven-Month-Old Infants and Neural Networks: Technical Comment by Altmann & Dienes and Response Marcus. Science, 284:875a.

- Aslin, R. N. (2007). What's in a look? Developmental Science, 10(1):48-53.
- Borensztajn, G., Zuidema, W., and Bechtel, W. (2014). Systematicity and the Need for Encapsulated Representations. In Calvo, P. and Symons, J., editors, The Architecture of Cognition: Rethinking Fodor & Pylyshyn's Systematicity Challenge, number 1988, pages 165-190. MIT Press.
- Cao, X., Sugiyama, Y., and Mitsui, Y. (1998). Application of artificial neural networks to load identication. Computers & structures, 69:63-78.
- Christiansen, M. H., Allen, J., and Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues : A Connectionist Model. Language and Cognitive Processes, 13(2/3):221-268.
- Christiansen, M. H. and Curtin, S. L. (1999). Transfer of learning: rule acquisition or statistical learning? Trends in cognitive sciences, 3(8):289-290.
- Dienes, Z., Altmann, G. T. M., and Gao, S.-j. (1999). Mapping across Domains Without Feedback: A Neural Network Model of Transfer of Implicit Knowledge. Cognitive science, 23(1):53-82.
- Elman, J. (1990). Finding structure in time. Cognitive Science, 14(2):179-211.
- Elman, J. and Seidenberg, M. (1999). Reply to marcus (science, april 17, 1999). Science.
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48(1):71-99.
- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28:3-71.
- Marcus, G. F. (1998). Rethinking Eliminative Connectionism. Cognitive Psychology, 37:243-282.
- Marcus, G. F. (1999a). Reply to Christiansen and Curtin. Trends in Cognitive Sciences, 3(8):290-291.
- Marcus, G. F. (1999b). Reply to Seidenberg and Elman. Trends in Cognitive Sciences, 3(8):289.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. Science (New York, N.Y.), 283(5398):77-80.
- Mpitsos, G. J. (2000). Attractors : Architects of Network Organization? Brain, Behavior and Evolution, 55:256-277.
- Negishi, M. (1999). Do infants learn grammar with algebra or statistics? Science, 284(5413):435.
- Plunkett, K. and Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. Cognition, 48(1):21-69.
- Seidenberg, M. S., Elman, J. L., Negishi, M., Eimas, P., and Marcus, G. (1999). Do infants learn grammar with algebra or statistics? Science, 284(5413):433-433.

Willits, J. A. (2013). Learning nonadjacent dependencies in thought,

language , and action : Not so hard after all. . . . In CogSci 2013, pages 1605-1610.

## Abstracts



#### Perceptual concordance of timbre similarity in electronic dance music

David Iñaki López Mejía

Supervisors: dr. Aline Honingh and dr. Makiko Sadakata

Centre for Digital Humanities, Institute for Logic, Language and Computation, University of Amsterdam

Most people can detect when two songs are similar to one another, but music similarity is yet to be defined as a quantitative construct. Previous studies have analyzed music similarity as a whole, but we discuss that it is, in fact, a multidimensional concept. For example, a music piece might be similar to another regarding the timbre, yet their rhythmic patterns could be rather different. In this case, these pieces of music would have high timbre similarity, but low rhythm similarity. In this study, we aimed to find 1) whether there is consensus among participants on the concept of timbre similarity for electronic dance music (EDM), 2) if there is within-participant consistency on the responses of an individual subject listening for timbre similarity, and 3) how far the concept of timbre similarity overlaps with general similarity for EDM. Our findings show that, although it seems possible to separate timbre from other dimensions of a musical segment and give ratings based on it with fair across-participant concordance and substantial within participant concordance, we cannot support the claim that timbre is always the most salient dimension when assessing general similarity.

#### Cross-situational associative word-learning in Dutch infants

Siméon T.A. Lahaije

Supervisor: dr. Caroline M.M. Junge

Department of Developmental Psychology, University of Amsterdam

It is a difficult task for infants to learn which word belongs to which referent due to the complexity of natural environments. However, it has been shown that infants are able to learn words in a cross-situational associative word-learning paradigm. To gain more insight in this learning process, we set up an experiment in which infants learned the non-sense names of six objects in a cross-situational paradigm with the words presented in natural utterances to facilitate learning. Results show that participants were able to learn words in this paradigm, and that they improved their scores over the course of the experiment, which suggests that they continue to learn while the experiment continues. These results are in line with earlier findings of cross-situational word-learning in infants. By understanding more about word-learning in general, it would be possible to detect developmental problems and treat them accordingly.

#### Individual differences in native speakers: comprehension of sentences with subject-verb broken agreement

Natalia Rivera

Supervisor: Sible Andringa

Amsterdam Center for Language and Communication, University of Amsterdam

The present study assessed Spanish native speakers' processing of sentences involving a subject-verb broken agreement. The purpose was to evaluate to what extent two distinct groups of participants differed in their grammatical processing, and ultimately in their linguistic competence in a reading comprehension experiment. Subjects were divided according to their educational attainment. Namely, whether they had university studies or not. It was hypothesized that subjects with low educational attainment (LAA) would perform a shallow parsing of the broken agreement, compared with participants with a high educational attainment (HAA), who would rely on a true, rule-based parsing strategy. Subjects' eye movements were measured in terms of first fixation on a critical region, first-pass reading times, total pass reading times, total reading times, and regressions to previous regions of the sentence. Results showed no significant effect of group in the first two measures, indicating that educational attainment, as a grouping variable, did not successfully predict participants' responses. As for the rest, main effects of condition were observed, with participants having longer fixations in those sentences that contained a real broken agreement. Regressions to regions that preceded the target region showed a preference from low academic achievement subjects for looking back at the noun causing the broken agreement. These results shed light on the discussion of how native speakers language processing might work: a shallow parsing, followed by a true-parsing heuristic might be taking place.

#### Modeling reward-guided decision-making with a biophysically plausible attractor network and the belief-dependent learning rule

Rebecca Sier

Supervisor: dr. Audrey Houillon

Neural Information Processing Group, Technische Universität Berlin

Elderly people and patients suffering from Parkinson's disease learn more from negative than from positive decision outcomes as compared to healthy controls. Previous research hypothesized that this error-avoidant behavior is partly caused by reduced levels of dopamine, and a model of this effect in the basal ganglia was constructed. However, the model is subject to discussion and hard to verify on the level of the basal ganglia. This thesis hypothesizes that the attractor model for decision-making combined with the belief-dependent learning rule can verify the basal ganglia model. It is shown how this combined model includes features of the basal ganglia model and explains them with greater detail. Simulations of the combined model make both correct and incorrect predictions of reward-guided decision behavior and neurotransmitter values of subjects that perform a probabilistic selection task. However, the empirical data is restricted and future research should be done to decide on the usefulness and the correct implementation of the model.

#### Trait Judgments: What do they reveal? Combining Trait Judgments With Real Personalities

Tineke Slotegraaf Supervisore: Andries van der Leij University of Amsterdam

We judge other people by their appearances faces, but are our judgements based on culture or on underlying traits of these people? Earlier research (Dotsch and Todorov, 2011; Engell et al, 2007; Oosterhof and Todorov, 2008) focussed on the judgments of faces, and showed that specific facial features lead consistently to trait inferences. However, since these researchers had no knowledge of the people on the pictures, it remains unclear if this relationship is adaptive. We tried to answer this question by letting 100 participants judge pictures of about 700 people who participated in the ID1000 project. This project was set up to see the differences in brains and behavior in a group representing the Dutch population. In the ID1000 project not only pictures were taken, they also collected data through questionnaires and a structural MRI scan. The judgments given to these people were correlated with the structural MRI scan using Voxel bases morphometry (VBM) for men and women seperately. Significant brain areas were extracted, the gray matter values were regressed against the Eigenfaces we generated from the pictures. The last step was to regress the brain area to the trait judgments to see which traits correlated with the brain structures and faces. For females the significant trait was extraversion, and for males the significant traits were dominance and openness.We did find a deeper cause for trait inference, in brain structure, but replication is needed to get better results and a better understanding of which facial features are used in trait inference.

#### Neural Mechanisms underlying the Development of Optimal Depth Cue Combination in childhood

Bauke van der Velde Supervisor: dr. Tessa Dekker University College London

The brain continuously receives information from the world through all sensory modalities. By integrating these cues in an optimal fashion, adults minimize variance to make more successful decisions. In children, however, optimal cue integration develops surprisingly late. This study explored two questions: At what age do children start integrating motion parallax and binocular disparity depth cues? And do these behavioural changes coincide with changes in the neural mechanisms underlying optimal cue integration? A behavioural experiment in which children were asked to determine which of two planes was further away, showed that only children of 11 years and older have started integrating motion parallax and binocular disparity depth cues in an optimal fashion. No evidence was found that suggests differences in neural mechanisms for cue integrators compared to non-cueintegrators. MVPA data of the fMRI experiment presented here thus supports the hypothesis that behaviour and neural mechanisms of optimal cue integration are dissociated during development. The delayed maturation of optimal cue integration might thus not be caused by young children lacking the tools to optimally integrate cues but by an inability to access this information efficiently at younger ages.

#### Cross-situational word learning in Dutch infants

Carola Werner

Supervisor: dr. Caroline M.M. Junge

Department of Developmental Psychology, University of Amsterdam

Word learning is difficult. One of the factors that could influence infants' success of word learning is auditory context; words may be presented either in isolation or in supportive carrier-phrases. Whereas parents tend to present novel words embedded in supportive, familiar sentence frames, experimental research studying novel word learning in the lab often present novel words in isolation. In the current study, we compared both modes of auditory context in a more ecologically valid word learning paradigm: cross-situational word learning. Within this paradigm infants have to keep track of object-word pairings in order to learn the correct mappings. To examine which auditory context promotes word learning better, the current study aimed to teach 18-month-olds six novel words which were presented in two different auditory contexts; isolation versus utterances. The results of this experiment showed that monolingual 18-month-olds learn equally well from both modes of presentation, although infants participating in the isolated words condition showed larger learning effects, which suggests that infants find words in isolation easier to learn.

7 questions about academic publishing

Dr. Judith E. Rispens Research interests:

language acquisition

developmental language disorders overlap between developmental dyslexia and SLI acquisition of written language morphophonology acquired language disorders

# 1. What, in your opinion, are the best and the worst aspects of the peer-review and publication process?

Best: Peer-reviews most often give you valuable feedback on your submission as they are performed by experts in the field. Worst: Peer reviews are not always completely objective as the person reviewing has his/ her opinions about the contents of the manuscript which are sometimes based on a different theoretical view.

### 2. What do you think could make the peer review and publication process better?

Setting firm time limits to the reviewers.

# 3. Do you feel that there is pressure to publish articles? Is this pressure appropriate for the advancement of science?

Yes, there is pressure, but to some extent people need to feel some pressure to perform well (at least, for me!). As long it is not publishing just for the sake of the number of publications, communication between researchers in the form of publications is an advancement. Conferences and conference papers are also important. It is a shame that null-results are difficult to publish, as knowledge of experiments of other researchers that do not render significant results is valuable and also leads to advancement of science.

#### 4. How long does it normally take for you to publish your research, from when you first start collecting data to when the paper is accepted?

Between 6 months and a year (or even 1,5 yrs) depending on the journal.

5. How often do you peer review manuscripts for journals? Do you feel that it is important to do peer reviewing, and why? Do you enjoy reviewing manuscripts?

Around 8 per year. It is important as you are involved in the publishing process, but it takes up time.

#### 6. Have you ever had to review a manuscript from a friend or colleague? Or from someone you do not like or respect? And do you think personal biases affect the review process?

I sometimes get invitations to peer-review work of colleagues, but I always reject these as I have done it once and I found it difficult to remain objective.

### 7. Do you have any advice for students who want their masters research to lead to a publication?

1) Ask a few people to comment on your draft before sending it to a journal

2) If you do experimental work, think carefully about administering tasks that provide 'baseline' info (depending on the theme e.g. IQ; handedness; vocabulary). Without this it is often difficult to publish your results.

3) And do not give up when you are rejected. Rewrite on the basis of the reviews and try again in a different journal.



# 7 questions about academic publishing

#### Jelle Bruineberg

Jelle Bruineberg is a Ph.D. candidate at the Institute for Language, Logic and Computation (ILLC) and the department of Philosophy of the Faculty of Humanities, at the UvA. He is part of the Erik Rietveld's VIDI-project 'Landscape of Affordances: Situating the Embodied Mind'. He is interested in how brain, body and world contribute jointly to everyday skillful behavior. Ecological psychology and phenomenology provide the right context in which to understand new developments in theoretical neuroscience and neurodynamics (such as the Free Energy Principle). For a first paper on this project, see Bruineberg and Rietveld 2014.



# 1. What, in your opinion, are the best and the worst aspects of the peer-review and publication process?

So far, I have relatively little experience with the peerreview and publication process, I only just published my first paper. In the good case, a reviewer can really make the paper better, while in the bad case a reviewer does not understand the paper properly and rejects it for the wrong reasons or insists that you clarify trivialities. Everyone I talk to tells me that in the end there is a factor of luck and randomness involved in who gets to review your paper and how they will take it. In an ideal case, you would of course want to factor out randomness, but this seems very hard to do.

### 2. What do you think could make the peer review and publication process better?

Some journal experiment with disclosing the reviewers after publication. I think this transparency could help people to take reviewing seriously. On the other hand, you might not want to openly and fiercely criticize your possible future employer. So, I have no clue really.

# 3. Do you feel that there is pressure to publish articles? Is this pressure appropriate for the advancement of science?

My work is situated in between the humanities and the sciences, and these two fields have very different publishing cultures. In philosophy, for instance, the PhD thesis is often not a collection of articles, but rather a whole book. One can publish parts of the thesis independently, but the focus lies on the thesis at the end of the PhD. Things are completely different in the sciences, where the aim of a PhD is to publish a number of studies and bundle them in the end. Also, in philosophy, papers are normally single-authored, while of course in the sciences the standard is to publish together with a great number of people. I try to find my way in between these different cultures.

In general, I do not feel a lot of pressure to publish a lot during my PhD, I think people in philosophy generally encourage you to take your time to really get into the debates and the literature. However, it is implicit that if you want to continue after the PhD, and this holds for both the humanities and the sciences, you need to be able to show you have been productive. In the end, published papers are still the standard measure to assess productivity.

#### 4. How long does it normally take for you to publish your research, from when you first start collecting data to when the paper is accepted?

In philosophy these things work a bit different than in the sciences. The paper that we have just published is the result of already a long collaboration. Me and my supervisor started thinking about writing a paper a bit more than a year ago and I have been writing on it for a bit less than a year next to my thesis work. I try to publish my Brain and Cognitive Science Master thesis as well, but this has still not happened almost a year after finishing my research project.

# 5. How often do you peer review manuscripts for journals? Do you feel that it is important to do peer reviewing, and why? Do you enjoy reviewing manuscripts?

I have not yet peer-reviewed manuscripts for journals, so I cannot say much about it. Although it seems to be quite time consuming, I think that I would enjoy peer-reviewing papers from other people. A healthy scientific community stands or falls with a critical and fair peer-reviewing process, so I think it is really important to do it.

#### 6. Have you ever had to review a manuscript from a friend or colleague? Or from someone you do not like or respect? And do you think personal biases affect the review process?

As said, I have not yet officially reviewed any papers myself, so I cannot say much about it, but at least in philosophy it is quite common to provide feedback on each other's papers before officially submitting to a journal. I have found that, at those instances, people are not biased to be friendly to each other, but rather aim to make the paper better even if that means to be harsh. In this sense, I do not think that peer-reviewing from friends induces a strong bias (although perhaps it does if they are the ones that need to decide whether to publish the paper or not).

### 7. Do you have any advice for students who want their masters research to lead to a publication?

I have only just finished my research master in Brain and Cognitive Science and my master in Philosophy (in January 2014), so what I can tell is from the perspective of a finished student. The research from my first internship did not get published (mainly because I showed that a particular model really can not be used to fit a particular kind of data, which is not a very publishable result), while the research of my second internship is still in preparation for publication. I did my second internship abroad, and during the first meeting with the head of the research group and my supervisor we discussed authorship for a possible publication. I think it is important to raise these questions early during the research project, even if your supervisor does not raise them: "do you plan to publish the research?", "how will we divide authorship?" If you want to publish a first-authored paper from your research project, it will be good to state from the beginning that this is your goal and ask what, if possible, the requirements for first-authorship are. Also, when in doubt, it is good to check what the norms for authorship are with people outside the research group, but within the field.

#### Berno Bucker

Berno Bucker is a Ph.D. candidate at the department of cognitive psychology at the Vrije Universiteit Amsterdam. He does behavioral, eye tracking and MRI experiments for Jan Theeuwes' ERC-project `What you get is what you see: How reward determines perception'. He is interested in how reward changes perceptual and attentional processes to guide adaptive behavior. His work is dedicated to both the motivational and the automatic priority effects that rewards elicit. For his first articles on reward see Bucker & Theeuwes (2014) and Bucker, Belopolsky & Theeuwes (2014).



# 1. What, in your opinion, are the best and the worst aspects of the peer-review and publication process?

What I like most about the peer-review process is that in the end your article is always incredibly improved. I had the privilege to receive seven peerreviews for three different manuscripts and almost all of the peer-reviewers had unique suggestions for improving those articles. At first sight, it might seem that a reviewer just wants to promote his or her own opinion or scientific work. Sometimes you might even think that he or she is just nitpicking on minor details to get your article blocked. But in the end, when all the control experiments are conducted and corrections are made, you have a look at the article and say "yes, this is it." For me, the worst aspect about the peerreview process is the waiting - sometimes several months between reviews. Once you have submitted the first version of your manuscript you are so excited that you cannot wait to receive the reviews, let alone when you sent the manuscript in for second or third revisions. When your manuscript finally gets accepted for publication, again, it might take two months for it to appear online.

### 2. What do you think could make the peer review and publication process better?

I believe that the interaction and communication with the authors and reviewers could be improved. On the one hand it is good that reviewers are anonymous, but on the other hand it feels old-fashioned to communicate by means of one review-letter and one response-letter. Sometimes it would be convenient to discuss issues that are brought up by the reviewers in person, by mail, phone or skype.

# 3. Do you feel that there is pressure to publish articles? Is this pressure appropriate for the advancement of science?

I think it depends a bit on the group you are working in, your professor and the number of fellow colleagues/ competitors. Personally I felt a lot of pressure at the start of my PhD but with every publication the pressure lessens. I think this is due to the system that we have in the Netherlands that requires you to write at least three to four articles to include in your dissertation. Personally I think this requirement sometimes hinders the learning process and puts too much pressure on PhD students to publish.

#### 4. How long does it normally take for you to publish your research, from when you first start collecting data to when the paper is accepted?

When I conduct behavioral research it is somewhat faster since multiple participants can be tested at the same time. This makes piloting a lot easier and faster, which speeds up the publication process. For eye tracking studies it takes somewhat longer and for EEG or MRI studies it will take even longer. For the behavioral and eye tracking experiments I think it is fair to estimate between 6-8 months for the whole process from data collection to manuscript acceptance.

# 5. How often do you peer review manuscripts for journals? Do you feel that it is important to do peer reviewing, and why? Do you enjoy reviewing manuscripts?

I have peer reviewed two manuscripts, which was a lot of work and quite challenging to do. At first you just try to criticize the paper, but then you realize that you are there to improve the work of someone else. Sometimes it can be really difficult to suggests the exact steps that need to be taken in order for the manuscript to improve. I kind of enjoyed peer reviewing, since you contribute to science in a different way compared to publishing your own articles. However, it is time consuming to write a good peer-review and you often do not get anything out of it directly. So in busy periods full of deadlines it can be quite stressful if you also have to peer-review the articles of others.

#### 6. Have you ever had to review a manuscript from a friend or colleague? Or from someone you do not like or respect? And do you think personal biases affect the review process?

I have never had to officially review a manuscript of someone I knew personally, but I am sure personal biases affect the review process. Even when you know the lab or the name of the author, I think you are already biased in your peer review.

### 7. Do you have any advice for students who want their masters research to lead to a publication?

If you are willing to go for a publication, I would suggest to thoroughly speak this through with your supervisor. They will immediately know if this is realistic, what needs to be improved and what steps need to be taken to start the publication process. Do not be shy and if you think it is possible to get a publication out of it, give it a try. However take into account that the whole process will take time and that you need to invest some (or a lot) of your free time in order to make it work.