Aberrant Hippocampal Morphology and Function in a Mouse Model for Post-Traumatic Stress Disorder

Leonie J. C. Madder[1]
Supervisors: Marloes Henckens[1], Judith Homberg[1], Tamas Kozicz[1,2]

[1]Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands
[2]Radboud University Medical Centre Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands

Post-Traumatic Stress Disorder (PTSD) is a debilitating disorder that affects approximately 8% of the population. It can be caused by experiencing a traumatic event and is characterised by symptoms like insomnia, hypervigilance and hyperarousal, and intrusive memories (i.e., flashbacks and nightmares) of the event which severely deteriorate quality of life. Overgeneralisation of the trauma memory has been suggested to underlie these intrusive memories triggered by indiscriminate environmental factors, and has been related to deficits in hippocampal function. Particularly pattern separation, the process by which memories are stored as unique representations resistant to confusion, mediated by the dentate gyrus (DG), may be compromised. However, since most evidence originates from patient studies that are done retrospectively, it is unknown whether this abnormal hippocampal functioning is in fact part of pathology or constitutes a predisposition to PTSD. Here, we used a validated mouse model for PTSD induction to induce a PTSD-like phenotype in part of the mice, whereas others are resilient and do not display any PTSD-related symptoms. This model perfectly mimics the human situation, in which only 20 - 25% of the individuals experiencing a trauma will ultimately develop the disorder, whereas the majority stays healthy. Using various methods, we assessed fear generalisation, and monitored hippocampal activity using arterial spin labelling, and the PTSD-associated neuroendocrine changes (i.e., corticosterone levels) over the course of PTSD development in PTSD-like compared to resilient mice. Moreover, Golgi staining enabled us to assess spine density in the ventral DG, to assess the potential for synaptic connectivity at the brain's site for pattern separation. In line with literature, PTSD-like animals displayed a suppressed corticosterone response to stress but no differential fear generalisation compared to resilient animals. Moreover, we observed reduced hippocampal activity in PTSD-like mice, but only after trauma induction, indicating the reflection of pathology rather than a predisposition. Lastly, spine density in the ventral DG was significantly reduced in PTSD-like mice. These results indicate that the hippocampal dysfunction associated with PTSD-like symptomatology in mice is a consequence of PTSD development, and that high stress-induced corticosterone levels might be protective for the development of PTSD.

Keywords: PTSD, fear generalization and pattern separation, ventral dentate gyrus, hippocampus, spine density, cerebral blood flow, corticosterone

Corresponding author: Leonie Madder; **E-mail:** leonie.madder@live.nl

Post-Traumatic Stress Disorder (PTSD) is a psychiatric disorder that affects approximately 8% of the population (Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). It can develop following exposure to a trauma that threatens the physical or psychological integrity of oneself or others (American Psychiatric Association, 2000). Symptoms that characterise the disorder are insomnia, hypervigilance and hyperarousal, flashbacks of the events and recurrence of memories of the trauma (Pitman et al., 2012), which can severely deteriorate one's quality of life. Its pathophysiological basis is poorly understood, making current treatments symptomatic and only effective for fewer than half of the patients. Hence, there is an urgent need to advance the understanding of its pathophysiology. Symptoms of flashbacks and recurrences of memories experienced by PTSD patients are thought to be caused by the process of fear generalisation (Greenberg, Carlson, Cha, Hajcak, & Mujica-Parodi, 2013), the process of generalising context-dependent fear to other (differing) contexts. Overgeneralisation of the trauma memory makes cues that are only remotely similar to those experienced during trauma capable of inducing a full-blown fear response (Golub, Mauch, Dahlhoff, & Wotjak, 2005; Greenberg et al., 2013), which could in turn relate to the typical avoidance symptoms observed in PTSD. Heightened fear generalisation has been linked to impairments in the hippocampus during the recollection of contextual fear memories in PTSD (Golub et al., 2005). In line with this, abnormal hippocampal functioning has been associated with PTSD (American Psychiatric Association, 2000; Bonne et al., 2001; Bonne et al., 2008; Shin, Rauch, & Pitman, 2006; Shin et al., 2004), with human magnetic resonance imaging (MRI) studies reporting on reduced hippocampal volumes and abnormal functioning such as memory deficits (Shin et al., 2004) in PTSD patients (Bonne et al., 2001; Shin et al., 2006). Animal models for PTSD have also reported abnormal hippocampal neuronal morphology after PTSD development. However, since these studies are usually done retrospectively (i.e., after PTSD diagnosis), it is unknown whether these abnormalities are the respective cause or consequence of the disease (American Psychiatric Association, 2000).

Fear generalisation is thought to originate in a particular subregion of the hippocampus, namely the dentate gyrus (DG; Ji & Maren, 2008). The DG is implicated in pattern separation, the process of extracting unique information from events that have overlapping or similar representations in order to make distinctions between internal representations of these similar memories (Aimone, Deng, & Gage, 2011; Sahay, Wilson, & Hen, 2011). Impairments in pattern separation have been related to an excessive generalisation of fear (Sahay et al., 2011), and it is thought to be an endophenotype for anxiety disorders like PTSD (Kheirbek, Klemenhagen, Sahay, & Hen, 2012). Granule cells are the primary cell type of the DG (Amaral, Scharfman, & Lavenex, 2007).These cells play a critical role in the hippocampal circuitry of health and disease, because these cells regulate the information flow into the hippocampus (Zohar et al., 2011) and thereby control contextual learning and features of anxiety (Kheirbek et al., 2013). The DG can be functionally segmented into the dorsal and ventral DG. The dorsal DG is primarily involved in information processing and cognitive functions needed for exploration, navigation, and locomotion, whereas the ventral DG is involved in regulating affect, motivational, and emotional behaviour and the hormonal stress response (Fanselow & Dong, 2010). In addition, the dorsal DG is implicated in encoding but not retrieval of contextual fear memories while the ventral DG is powerful in suppressing innate anxiety (Kheirbek et al., 2013). Also, abnormal function of the dentate gyrus in particular seems to be associated with PTSD (American Psychiatric Association, 2000; Kheirbek et al., 2012; Kheirbek et al., 2013; McEwan, 1999; Sahay et al., 2011; Zohar et al., 2011).

However, the DG is highly sensitive to stress (Kavushansky, Vouimba, Cohen, & Richter-Levin, 2006) and elevated corticosterone levels are especially harmful to the DG (McEwan, 1999). Also, a relationship between the density of spine in the dentate gyrus and stress is found in the literature and density changes are implicated in post-traumatic stress disorder (Adamec, Hebert, Blundell, & Mervis, 2012; Cohen, Kozlovsky, Matar, Zohar, & Kaplan, 2014; Diamond et al., 2006; Dias et al., 2014; Shors & Leuner, 2004; Zohar et al., 2011). Spines play a role in synaptic plasticity (Nimchinsky, Sabatini, & Svoboda, 2002), and are involved in the establishment and maintenance of connections with axons from other neurons (Bohlen & Halbach, 2009). Spines are important for neural processes, and changes in density might reflect adjustments in synaptic transmission (Bohlen & Halbach, 2009). These changes in spine density, and subsequently the changes in synaptic transmission are thought to underlie psychopathological and pathophysiological alterations (Bohlen & Halbach, 2009). Many studies have reported the existence of a relationship between stress and spine density in the dentate gyrus. However, this literature is often conflicting.

Dias et al. (2014) showed that spine density in the rat DG is increased in a model for generalised anxiety disorder (Dias et al., 2014), while others have shown a reduction in spine density in the rat DG in response to acute stress (Adamec et al., 2012; Cohen et al., 2014; Zohar et al., 2011). Additionally, PTSD is associated with alterations in the hypothalamic-pituitary-adrenal (HPA) axis, resulting in hypocortisolism by increased negative feedback (Van Zuiden et al., 2012). Although generally reduced cortisol secretion has been observed in PTSD patients (Yehuda, Teicher, Trestman, Levengood, & Siever, 1996), animal studies trying to find a causal link between suppressed corticosterone signalling and PTSD susceptibility are conflicting. In these studies, the effects of corticosterone administration following trauma (to increase the potentially suppressed corticosterone response contributing to PTSD development) are investigated. On the one hand, Zohar et al. (2011) found that the administration of a high dose corticosterone immediately after trauma exposure reduced the risk of PTSD development, and prevented the trauma-induced reduction in dorsal DG spine density as observed in the extreme responding animals. However, on the other hand, Kaouane et al. (2012) found that corticosterone, when administered following severe stress, induced PTSD-like fear memory generalisation, and increased fear responses when administered after relatively mild stress. Because of these contractor findings, the exact contribution of corticosterone to PTSD development and DG neuronal morphology remain elusive and further research is needed to elucidate its exact role in PTSD. Here, in a prospective study, we wanted to elucidate the causal relationship between neuroendocrine parameters (i.e., corticosterone levels), fear generalisation, and hippocampal function and the susceptibility to PTSD. Therefore, mice are exposed to a validated mouse model for the induction of PTSD (Lebow et al., 2012), in which mice are first exposed to a severe stressor (i.e., intense electric foot shock) in a certain context (context A), followed by a mild stressor (i.e., weak electric foot shock) the next day in a distinct context (context B) which is different in terms of spatial, auditory, and olfactory cues. By stressing the animals in two distinct contexts, this protocol actively stimulates the process of fear generalisation (of context A to B), and has been shown to reliably induce PTSD-like symptomatology (i.e., hypervigilance, compulsivity, impaired attention, compromised risk assessment, and insomnia) in a subset of mice, whereas others do not show any of these behaviours and are resilient. This observation perfectly mimics the human situation in which only a relatively small fraction (15 - 20%) of the individuals exposed to a traumatic event develops PTSD (Breslau, 2001), but is typically ignored in animal studies implementing very severe traumas to (artificially) warrant PTSD development (Servatius, Ottenweller, & Natelson, 1995). It is, however, unknown what causes this differential vulnerability, and an increased understanding would open up new avenues for treatment. Fear generalisation was measured by the amount of freezing during re-exposure of a traumatic experience (i.e., during trigger exposure). Basal corticosterone levels were measured both at the start and at the end of the protocol, as well as trauma- and trigger-induced corticosterone stress response and corticosterone responses to restraint stress after the development of pathology. To monitor hippocampal function, we obtained two MRI scans (i.e., arterial spin labelling [ASL]; Zerbi et al., 2014), before and after PTSD induction, to identify whether changes in hippocampal activity might relate to PTSD development. Following PTSD induction, we analysed spine density in the ventral DG of PTSD-like and resilient mice to investigate potential differences in the DG morphology following PTSD development. We hypothesised that PTSD-like animals would display reduced hippocampal activity, suppressed corticosterone stress responses and a reduced ventral DG dendritic spine density following PTSD induction compared to their resilient counterparts. The aim of the study was to elucidate whether these factors indicate possible vulnerability to PTSD or rather pathology as a consequence of PTSD.

Methods

Animal background and maintenance

92 C57BL/6J mice were housed in groups of four animals per cage with ad libitum access to food and water in a pathogen-free, temperature-controlled (22 °C ± 1) room with a humidity of 50 - 52%, and on a reversed 12/12 h light/dark cycle (lights on at 20:00 h). The animals were housed in the Central Animal Laboratory animal facility (Centraal Dierenlaboratorium [CDL], Nijmegen, the Netherlands), where the experiments took place. The study consisted of two separate experiments, experiment 1 (n = 32) to assess ventral DG spine density, and experiment 2 (n = 60) for the monitoring of hippocampal activity using

MRI scanning. Different group sizes were chosen based on power analyses on the distinct read-out measures of the experiments (morphology vs. blood flow), characterised by differential sensitivity and variability. As the proposed PTSD model was only validated in male mice, and stress responses in females differ from those in males and are related to the phase of the oestrous cycle (Ter Horst, De Kloet, Schachinger, & Oitzl, 2012), we only include male mice. At the start of the experiment, the animals were 10 weeks old. The experimental protocol was approved by the Animal Experiment Committee Radboud University (Dierexperimenten commissie Radboud Universiteit Nijmegen; DEC) in Nijmegen, and it was in accordance with the guidelines for the Care and Use of Mammals in neuroscience and Behavioural Research (National Research Council 2003), the principles of laboratory animal care, as well as the Dutch law concerning animal welfare. All experiments were carried out in accordance with the guidelines of the European Communities Council Directive.

Experimental design

See Figure 1 below.

PTSD induction protocol

PTSD induction. The PTSD induction protocol used in the experiment is based on the previously validated mouse model for PTSD, established by Lebow et al. (2012), to induce a PTSD-like phenotype in a subset of susceptible mice. In this model, on the first day of PTSD induction (day 15/16 in the experiment), mice receive 14 shocks of 1 mA, 1 s in duration with a continuous pulse over 85 minutes in variable intervals, representing the "trauma". For trauma induction, mice were placed in a fear conditioning apparatus (TSE Systems, Bad Homburg, Germany), using a certain context A. This context consisted of a transparent, Plexiglas cage (21 × 20 × 36 cm) with a metal grid floor and illumination of 10 lux with 70 dB background noise generated by the ventilation system. Inside the cage, a white paper was placed under the metal grid and a semi-circular white opaque screen was placed in the back of the cage to create a white, rounded environment. The mice were individually transferred to the experimental room in fresh cages in the light and the lights in the experimental room were kept on. Between experiments, the cage and grid were cleaned with 70% ethanol solution and sprayed with 1% acetic acid solution to create a scent. On the second day (day 16/17 in the experiment), the same mice received another 5 shocks of 0.7 mA, 1 s in duration over 5 minutes in fixed intervals, representing the "trigger", in a different context B. Context B existed of a black, solid cage (21 × 20 × 36 cm) with a triangular black screen placed over the cage to create a dark environment. There was no illumination inside the box and the experimental room except for infrared light. The mice were transferred to the experimental room in small dark carton cages in pairs. Between sessions, the cage and the metal grid were cleaned with 70% ethanol solution and additionally sprayed with 70% ethanol solution to create a scent.

Freezing behaviour. Video recordings made during trauma and trigger exposure were analysed for freezing behaviour of the animals by The Observer 5.0 (Noldus, Wageningen, The Netherlands). Freezing behaviour was scored by an observer, blinded to the experimental groups. Freezing was scored as the complete immobility of the animal for more than 1 second. To screen for potential

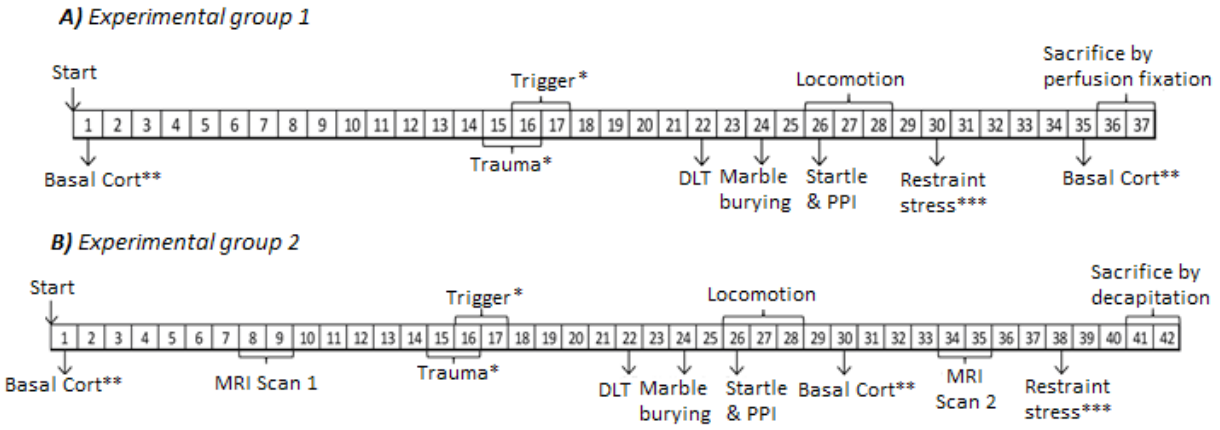

Fig. 1. Experimental design experimental group 1 and 2. DLT = dark/light transfer test, PPI = pre-pulse inhibition, Startle = startle response, CORT = corticosterone. *10 µl blood samples taken of the animals.

differences in both the initial response to the trauma, as well as its ultimate behavioural effect, the duration of freezing during trauma exposure was measured during the first 5 minutes (at t = 3 min, the first shock occurred and the second shock at t = 5 min) and the last 5 minutes of trauma exposure (at t = 82 min, the fourteenth [last] shock occurred, after which a 4 minute interval without shocks was presented until the end of the trauma session).

Behavioural assessment. All behavioural testing was performed during the dark phase, which is the active phase of the animals. The testing was performed at least four hours after the lights were turned off (from 12:00 h onwards) to ensure stable and low corticosterone plasma levels during testing. Following the PTSD induction protocol, the mice were subjected to five behavioural tests: the dark/light transfer test to assess risk taking behaviour (on day 22), the marble burying test to assess hypervigilance (day 24), acoustic startle response and pre-pulse inhibition to assess hyperarousal (day 26), and total light activity as an index for insomnia (days 26 - 28). The experimental setups were cleaned with 70% ethanol between every testing session. All animals from the same cage were tested consecutively and only after all animals from the same cage were tested, the mice were reunited in their home cage and returned to their housing room.

Risk assessment. In order to measure risk assessment, the dark/light transfer test was conducted. The test took place in a box with two compartments, with a partition in between. The dark compartment (15 × 25 × 25 cm) was covered with a dark plate and the light compartment (30 × 25 × 25 cm) was brightly illuminated by 1000 – 1100 lux. A small passage (3 × 6 cm) at the bottom of the partition connected the two compartments with a retractable door that opened the passage to initiate the test. At the start of the 5 minute test session, the mouse was placed in the dark compartment, and recordings of the test were made by a small camera installed above the apparatus. Data were analysed by Ethovision 10.1 xt (Noldus, Wageningen, The Netherlands). A risk assessment zone surrounding the opening of the passage in the light zone (3 × 6 cm) was programmed into the software tracking measurements.

The latency to enter the light compartment, the number of visits and the time spent in the risk assessment area were analysed. Percentage risk assessment time was calculated as the amount of time spent in the risk assessment arena as a percentage of total time spent in the light area outside of the risk assessment zone.

Marble burying. The mice were placed in a black box (30 × 27 × 26 cm), without a roof, with an illumination of 10 lux. The box contained 5 cm autoclaved bedding, and on top of the bedding 20 marbles were centrally arranged 4 by 5. The mice were filmed for 25 minutes by a camera mounted above the box. The videos were manually scored by counting the number of buried marbles at 25 minutes (Lebow et al., 2012).

Acoustic startle and pre-pulse inhibition. The acoustic startle response (ASR) (StartleResponse, TSE-systems, Bad Homburg, Germany) protocol was adapted from Lebow et al. (2012). A mouse was placed in a small Plexiglas and wire mesh cage in a sound-attenuated, ventilated chamber.

The cage was placed on top of a vibration-sensitive platform with a high-precision sensor detecting movements. Audio stimuli inside the chamber were produced by two high-frequency loudspeakers. 70 dB white background noise was maintained throughout the whole test. The ASR session started with a 5 minute acclimatisation period. After this acclimatisation session, randomly selected twenty-four startle stimuli of 120 dB, 40 ms in duration with a variable ITI of 12 – 30 ms were executed. Alternated with these startle stimuli, thirty-six additional startle stimuli with randomly preceding pre-pulses (20 ms) of either 75 dB, 80 dB, or 85 dB were applied. Maximal ASR, latency to peak startle amplitude, and the amount of pre-pulse inhibition (PPI) were individually analysed following the ASR protocol. The latency to peak startle was defined as the time from onset of the startle sound until the peak amplitude of the first startle. The time of latency to peak startle was averaged over all startle response trials. The percentage PPI was defined as the percent difference of the maximal ASR to the startle stimuli preceded by pre-pulses compared to the startle stimuli without pre-pulses, which was then averaged over all pre-pulse intensities.

Day-night locomotion. Mice were individually housed for 72 hours in Noldus Phenotyper 4500 (Noldus, Wageningen, The Netherlands) consisting of Plexiglas cages (45 × 45 × 60 cm) with a camera mounted on top. The first 24 hours was considered as a habituation period to the individual housing condition, after which the animals were tracked for 48 hours, consisting of two light and two dark cycles. The locomotion measurements were collected in 10 minute intervals and

analysed for total light and total dark locomotion by the Noldus software Ethovision 10.1 xt.

PTSD categorisation. The aforementioned behavioural tests were attributed to a particular amount of points according to factor analysis established by Lebow et al. (2012). The tests were divided into three groups: 1) percentage risk assessment and latency to peak startle amplitude, 2) percentage PPI, and 3) total light activity and marble burying. This resulted in the following scoring of extreme behavioural outcomes: Top 20% of animals showing high marble burying and high total light activity were attributed 1 point, bottom 20% of animals showing extremely low risk assessment were attributed 3 points, bottom 20% showing extreme latencies to peak startle were attributed 3 points, bottom 20% of abnormally low PPI were attributed 2 points (see Table 1). The scores per animal were tallied. This resulted in a behavioural phenotype that could be used to classify the animals in resilient or PTSD-like. Animals were considered resilient when they obtained a total score of 0 (and thus displayed no extreme behaviour in any of the tests), and as PTSD-like when the total score was 4 or more (reflecting extreme behaviour on multiple tests).

Table 1.

Inclusion criteria PTSD-like and resilient phenotype.

Behavioural measure		**Score**
Marble buying	Top 20%	1
Total locomotion light phase		1
Latency to peak startle		3
% Prepulse inhibition	Bottom 20%	2
% Risk assessment		3
	PTSD-like	≥ 4
	Resilient	<1

Neuroendocrine assessment

Blood sample collection during the experiment. To assess potential differences in corticosterone levels between PTSD-like and resilient mice, 10 µl blood samples were collected on various moment during the experiment. Blood sampling through tail bleed is in itself only mildly aversive to the animal, as it is done rather quickly (typically < 30 s) and does not entail any further restraint. Therefore, it can be conducted multiple times without negative consequences. Before PTSD induction, baseline peak and trough diurnal levels were assessed by two blood samples taken at 7:30 h (30 minutes before the start of the active phase to capture the circadian peak) and at 20:30 h (30 minutes after the end of the active phase to assess the circadian trough). Blood samples were also collected immediately after trauma and 10 min post trigger exposure. Three more samples were taken to assess the corticosterone stress response to restraint after PTSD induction (day 30 for group 1, day 38 for group 2). During restraint, animals were put in plastic restrainers for 25 minutes in a brightly lit room. Blood samples were collected by tail bleed before (t = 0 min, baseline), immediately after (t = 25 min, stress-induced peak levels) and 90 minutes after stress initiation (following recovery). Finally, towards the end of the experiment, another two basal levels were taken to assess the circadian peak and trough levels.

Corticosterone collection and measurement. 10 µl blood samples of were collected by tail bleed. After collection, blood samples were briefly stored on ice until they were centrifuged (3500 rpm for 20 min at 4 °C), and plasma was extracted. The plasma samples were stored at -20 °C and later assayed for corticosterone using the Corticosterone Double Antibody RIA Kit (MP Biomedicals, Orangeburg, NY, USA).

MRI procedure. All animals from the second experiment group, n = 60, underwent two MRI sessions, before and after PTSD induction protocol. The mice were anaesthetised with 3.5% isoflurane and transferred to the MRI platform, where anaesthesia levels of isoflurane were reduced to 2%, administered by inhalation through a nosetube. The mice subsequently received a bolus of medetomidine (Dexdomitor, Pfizer, 0.05 mg/kg) subcutaneously (Grandjean, Schroeter, Batata, & Rudin, 2014). After five minutes, the isoflurane was further reduced to 1%, and another five minutes later it was lowered to 0.5%, and infusion of medetomidine (0.1 mg/kg/h) started (Grandjean et al., 2014), which was maintained throughout the scanning session to maintain the superficial sedation level.

The mice were placed in a MR-compatible stereotactic device and immobilised with earplugs and a tooth holder. Body temperature was measured using a rectal thermometer and maintained at 37 °C using a heated air flow device. The respiration rate

was registered using a breathing pad. To protect the eyes from dehydration, eye ointment was used.

The animals underwent approximately 2 hours of MRI scanning. Two ASL scans to assess the cerebral blood flow (CBF) in the hippocampus were acquired at 45 to 60 minutes after the medetomidine bolus. After completion of scanning, the animals were removed from the apparatus, halting the administration of isoflurane and the medetomidine infusion, and a bolus (0.25 mg/kg) of antisedan (Atipamezole, Pfizer) was administered subcutaneously to antagonise the medetomidine and ensure quick recovery (Adamczak, Farr, Seehafer, Kalthoff, & Hoehn, 2010).

MRI acquisition. MRI measurements were performed on an 11.7 T BioSpec Avance III small animal MR system (Bruker BioSpin) equipped with an actively shielded gradient set of 600 mT/m and operated by Paravision 6.0 software. A circular polarised volume resonator was used for signal transmission and an actively decoupled mouse brain quadrature surface coil was used for signal reception (Bruker BioSpin).

To study brain perfusion under resting conditions, a flow-sensitive alternating inversion recovery arterial spin labelling (FAIR ASL) technique was used (Zerbi et al., 2014; Field of View = 25 × 25 mm; matrix size = 512 × 512; slice thickness = 1 mm; repetition time [TR]/echo time [TE] = 12 s/10.077 ms; spatial resolution = 0.195 × 0.260 × 1 mm/pixel). Sixteen images with increasing inversion times (TIs; 100 – 1600 ms) were obtained for the T1 calculations, adding up to a total scan time of 12 minutes 48 seconds. Inversion recovery data from the imaging slice were acquired after sequential selective inversion and nonselective inversion. Slices were positioned based on Bregma coordinates; centered around Bregma = -1.94 mm for the hippocampus (Allen Mouse Brain Atlas).

MRI analyses. For each mouse, the FAIR images with different TIs were realigned over the first TI. For this, a rigid-body model was used, implemented in Statistical Parametric Mapping (SPM). Determination of $T1_{selective}$ and $T1_{nonselective}$ was performed by fitting the averaged signal intensities in each region of interest (ROI) with a three-parameter monoexponential T1 relaxation curve. Using the following equation, CBF was determined in hippocampus:

$$\frac{CBF}{\lambda} = \frac{T1_{nonselective}}{T1_{blood}}\left(\frac{1}{T1_{selective}} - \frac{1}{T1_{nonselective}}\right)$$

In this equation, λ is the blood/tissue partition coefficient for water (assumed to be 0.9 ml/g) and T1blood was assumed to be 2.75 s at 11.7 T (Zerbi et al., 2014).

The regional cerebral blood flow was determined by drawing regions of interest in the retrieved ASL MRI scans in MatLab. The MRI procedure was performed by two independent researchers, who were blinded to the experimental conditions.

Brain tissue collection

Mice from experimental group 1 (n = 32) were sacrificed 20 days after trauma exposure by perfusion with 4% paraformaldehyde (PFA) in 0.1 M PBS. The brains were removed immediately after death and post-fixed for 24 hours by PFA immersion. After post-fixation, the brains were stored at 4°C. At a later stage, Golgi staining was performed on the brains.

Mice from experimental group 2 (n = 60) were sacrificed 25 days after trauma exposure by decapitation. Immediately after death, brains were removed, quick-frozen on dry ice and stored at -80 °C.

Golgi staining

Tissue preparation. The brains of animals from experimental group 1 that had undergone the PTSD induction protocol and were scored with either PTSD-like (n = 5) or resilient (n = 6) phenotype, were processed for rapid Golgi-Cox staining (the brains of mice with an intermediate phenotype were not included). The Golgi-Cox-based kit (FD Rapid GolgiStain™ FDNeurotechnologies, Inc. Ellicott City, MD, USA) was used for Golgi-Cox tissue impregnation. Brain tissues were immersed in a solution consisting of a mixture of mercury chloride, potassium dichromate and potassium chromate, and kept in the dark for 14 days. After impregnation, the tissues were transferred into a cryoprotectant solution for 72 hours to equilibrate and sink. After this, the tissues were cut into 140 µm coronal sections using a freezing-stage sledge microtome (Microtom HM440E, GMI Inc., Ramsey, MN, USA), and mounted on gelatine-coated slides. The sections were left to air-dry for 3 days before the sections were washed twice for 4 minutes in dH_2O, and incubated in an ammonium hydroxide solution for ten minutes. Then, the sections were washed twice in dH_2O for 4 minutes, placed for 4 minutes in ascending grades of alcohol (70%, 95% and four times 100%) and placed three times for 4 minutes each in

xylene for clearance. As a last step, the sections were coverslipped with Permount™ mounting medium (Fisher Chemicals, Leicestershire, UK) and dried for 1 week until the sections were processed for analysis.

Spine density. Prior to the quantitative analysis of neuronal spine density, a set of criteria was determined in order to obtain accurate measurement of neuronal morphology. Neurons were included into analyses if they fulfilled the following selection criteria: neurons should be 1) granule cells obtained from the granule cell layer of the ventral dentate gyrus (-3.28 to -3.80 mm from Bregma) in both the right and left hemisphere 2) completely impregnated, and 3) granule cells containing a clearly stained cell body and primary dendritic branch that was distinguishable from neighbouring neurons. Spine counting was performed on tertiary (3), quaternary (4), quinary (5), senary (6) and septenary (7) order dendrites.

Spine density measures were obtained using a Zeiss Axioskop FS microscope (Oberkochen, Germany). Dendritic spines were captured using a 100x oil immersion objective lens (plan-NEOFLUAR, Zeiss). The dendritic spines were traced using Neurolucida 11.09 (MBF Bioscience, Microbrightfield Inc. Williston, VT, USA) by an investigator that was blind to the treatment conditions. Firstly, the whole apical dendritic tree was traced. Secondly, the spines were counted per selected segment of at least 50 µm in length. Spine density was calculated as the number of spines per 10 µm of dendrite. On average, six segments per cell and five cells per animal were used for analysis. It should be noted that the spine density values obtained are likely to be an underestimation of the actual density of the dendritic spines because spines protruding beneath or above the dendritic segment were not visible and thus not accounted for.

Statistical analysis

Statistical analyses were computed with IBM SPSS statistics 23 (IBM corporation, Armonk, NY, USA). Group differences between PTSD-like and resilient animals in terms of behavioural measures, corticosterone levels and spine density were tested for using independent-samples t-tests. Hippocampal CBF was analysed using a repeated measures ANOVA, followed up by paired and independent t-tests. Spearman's rho correlations were performed for correlational analyses. Results were considered significant if $p < .05$.

Results

Behavioural assessment of PTSD-like phenotype in mice

Experimental group 1. As expected, PTSD-like animals showed impaired risk assessment during the dark/light transfer test ($t(9) = -2.78$, $p < .05$; Fig. 2A), and attenuated pre-pulse inhibition ($t(9) = -2.23$, $p = .05$; Fig. 2C). However, differences in marble burying behaviour ($t(9) = 0.92$, $p = .38$; Fig. 2E), latency to peak startle amplitude ($t(9) = -0.27$, $p = .79$; Fig. 2B), and activity during the light phase, ($t(9) = 1.47$, $p = .17$; Fig. 2D) failed to reach significance.

Experimental group 2. As expected, PTSD-like animals showed impaired risk assessment during the dark/light transfer test ($t(2) = -2.17$, $p < .05$; Fig. 3A), had a reduced latency to peak startle response ($t(22) = -3.45$, $p < .01$; Fig. 3B), and buried significantly more marbles ($t(22) = -2.83$, $p < .01$; Fig. 3E). However, PTSD-like mice did not significantly differ for the resilient ones in terms of pre pulse inhibition ($t(22) = -0.95$, $p = .34$; Fig. 3C), and their increased activity during the light phase just failed to reach significance ($t(22) = 1.78$, $p = .09$; Fig. 3D).

Behavioural response during the PTSD induction protocol: Trigger exposure

Experimental group 1. Freezing behaviour during trigger exposure was measured to assess fear generalisation. There was no significant difference in latency to freeze ($t(5) = 1.69$, $p = .15$; Fig. 4A) between PTSD-like and resilient mice. Also the duration of freezing was not significantly different between the two groups ($F(1,6) = 0.52$, $p = .50$; Fig. 4B). Both innate fear generalisation, prior to shock exposure ($t(9) = -0.82$, $p = .44$) and triggered fear generalisation, after first shock ($F(1,6) = 0.52$, $p = .45$) were not significantly different between the two groups, although the freezing duration over time was significantly increased ($F(1,6) = 173.58$, $p < .001$; Fig. 4B).

Experimental group 2. PTSD-like animals had a significantly shorter latency to freeze than resilient animals ($t(20) = -3.31$, $p < .005$; Fig. 5A). The duration of freezing did not significantly differ between the groups ($F(1,21) = 0.2$, $p = .90$; Fig. 5B). Both innate fear generalisation, prior to shock exposure ($t(22) = 1.64$, $p = .12$) and triggered fear generalisation, after first shock ($F(1,22) = 0.21$, $p = .65$) were not

significantly different between the two groups. However, there was a significant increased freezing effect over time ($F(1,21) = 350.03, p < .001$; Fig. 5B).

HPA-alterations in PTSD-like versus resilient mice

Experimental group 1. PTSD-like mice displayed a significantly increased pre-exposure circadian peak corticosterone level ($t(6) = 5.08, p < .005$; Fig. 6A), although circadian trough level was not significantly different between the groups ($t(9) = -1.49$, $p = .17$; Fig. 6A). The post-exposure circadian peak ($t(9) = -0.50, p = .63$; Fig. 6B) and trough ($t(9) = -0.53$, $p = .61$; Fig. 6B) levels were not significantly different between the groups. As expected, the PTSD-like mice had a significantly suppressed corticosterone level during restraint stress at $t = 25$ min ($t(8) = -3.14$, $p < .05$; Fig. 6D) and trend-level significant increase during restraint stress baseline measures ($t(8) = -2.13$, $p = .06$; Fig. 6B), which replicates the results of Lebow et al. (2012). Interestingly, corticosterone levels after trigger exposure ($t(9) = -2.91, p < .05$; Fig. 6C) are also suppressed in PTSD-like mice.

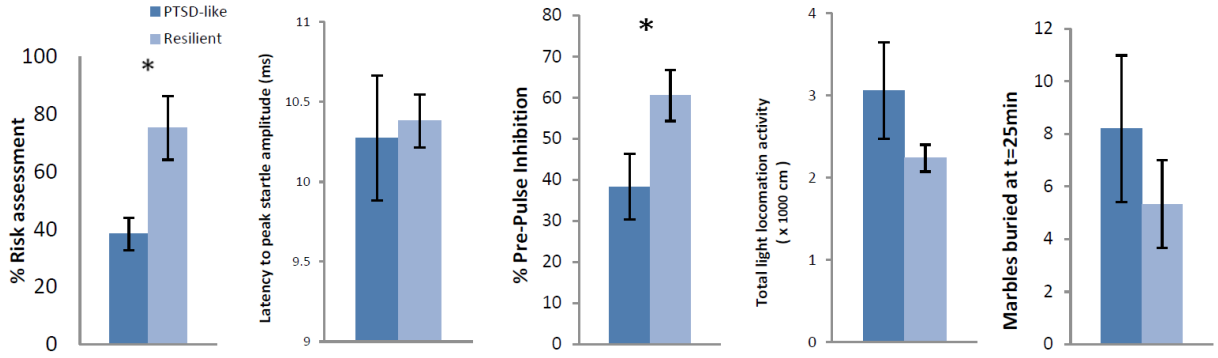

Fig. 2. Behavioural assessment. PTSD-like mice spent significantly less time engaged in risk assessment behaviour ($p = .02$) compared to resilient mice. No significant differences were detected ($p = .79$) for the latency to peak startle between the groups. Pre-pulse inhibition was significantly impaired ($p = .05$) in PTSD-like mice compared to resilient. PTSD-like mice tended to be more active in the light phase compared to resilient mice, but this difference failed to reach significance ($p = .18$). PTSD-like mice did not bury significantly more marbles ($p = .38$) then resilient mice. *$p < 0.05$. Error bars indicate standard errors of the mean (*SEM*).

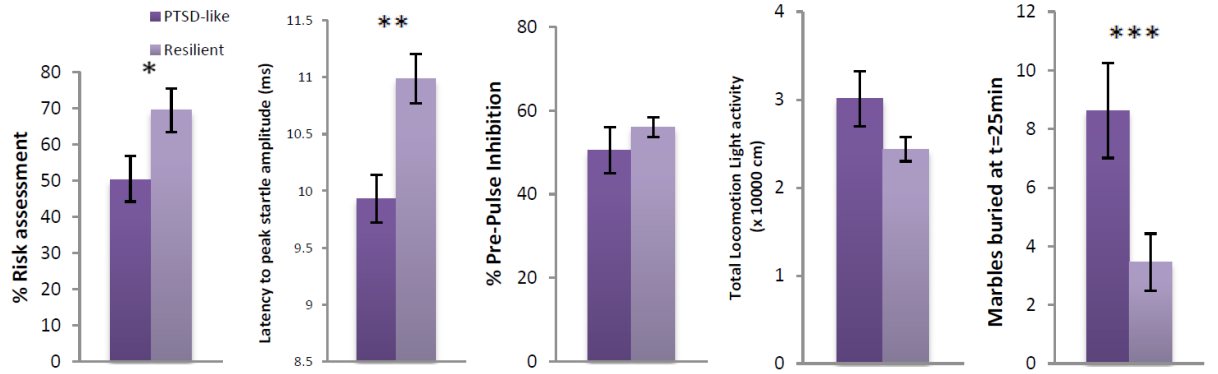

Fig. 3. Application of inclusion criteria for group 2. PTSD-like mice spent significantly less time engaged in risk assessment behavior ($p = .04$) compared to resilient mice. Significantly reduced latency to peak startle for PTSD-like animals ($p = .002$) was found. Pre-pulse inhibition was not significantly impaired ($p = .35$) in PTSD-like mice compared to resilient. PTSD-like mice were significantly not more active in the light phase ($p = .09$) compared to resilient mice. PTSD-like mice buried significantly more marbles ($p = .01$) compared to resilient mice. *$p < .05$, **$p < .01$. Error bars indicate *SEM*.

Experimental group 2. No significant differences between the groups were found during pre-exposure circadian peak ($t(21) = 1.42$, $p = .17$; Fig. 7A) and trough ($t(20) = 0.33$, $p = .75$; Fig. 7A) levels, and post-exposure circadian ($t(22) = -0.64$, $p = .53$; Fig. 7B) peak and trough ($t(21) = 0.26$, $p = .79$; Fig. 7B) levels. PTSD-like mice had a significantly increased baseline corticosterone level during restraint stress ($t(20) = 3.02$, $p < .01$; Fig. 7D). However, resilient mice had a significantly increased corticosterone level during recovery of restraint stress ($t(21) = -1.93$, $p = .07$; Fig. 7D).

Ventral DG spine density in PTSD-like versus resilient mice

Spine density analysis on the ventral DG revealed a significant reduction in spine density in PTSD-like animals ($t(9) = -2.22$, $p < .05$; Fig. 8C) compared to resilient animals (Fig. 8).

Hippocampal blood flow before and after PTSD induction protocol in PTSD-like versus resilient mice

Prior to PTSD induction, no significant differences ($t(20) = 0.96$, $p = .89$; Fig. 9C) were detected between PTSD-like and resilient mice. However, after PTSD induction, the hippocampal blood flow was significantly reduced ($t(18) = -2.09$, $p < .05$; Fig. 9C) in PTSD-like mice compared to resilient mice. A trend-level significant time x group interaction in hippocampal CBF levels ($F(1,17) = 3.30$, $p = .09$) was observed, without an overall effect of time ($F(1,17) = 2.29$, $p = .15$). Post-hoc analysis revealed that this group x time interaction was caused by a relative decrease in hippocampal blood flow in the PTSD-like group only ($t(6) = 2.28$, $p = .06$; Fig. 9C).

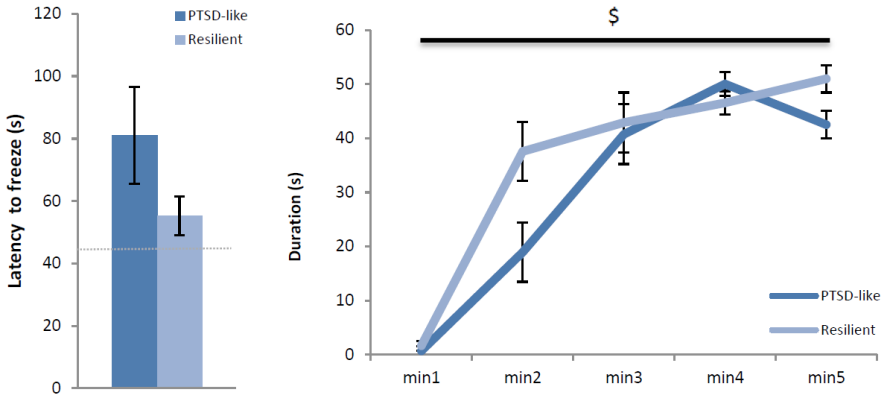

Fig. 4. Freezing behaviour during trigger exposure in mice from experiment 1. No significant difference was found in latency to freeze. Dashed line indicates the timing of the first shock (after 60 seconds). No significant difference was detected in the freezing duration between PTSD-like and resilient mice. However, there was a significant freezing effect over time ($p < .001$). Error bars indicate *SEM*.

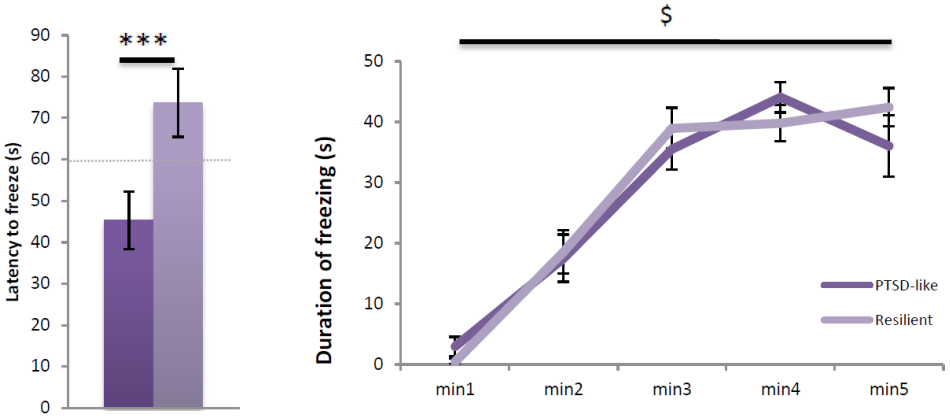

Fig. 5. Freezing behaviour during trigger exposure in mice from experiment 2. Resilient mice had an increased latency to freeze ($p = .003$) compared to PTSD-like mice. There are no differences in freezing duration ($p = .90$), although there is a significant increase in freezing over time ($p < .001$). ***$p < .005$. Error bars indicate *SEM*.

Correlational analyses

To gain additional understanding about the relationship between corticosterone levels, freezing behaviour, spine density and hippocampal blood flow, correlations were computed between measures.

A positive correlation was found between PTSD-like mice from group 1 for the spine density and post-PTSD development basal morning corticosterone levels ($r(5) = 0.90$, $p < .05$), and between trigger freezing latency and post-PTSD development basal morning corticosterone levels ($r(5) = 0.90$, $p < .05$).

Discussion

Symptoms of PTSD, like flashbacks and recurrence of memories, are thought to be caused by abnormal hippocampal functioning, and more specifically by exaggerated fear generalisation. Here, we investigated the contribution of hippocampal function and fear generalisation to PTSD. We particularly aimed to address the existence of potential differences in hippocampal function reflecting vulnerability for PTSD predicting its development. We used a validated mouse model for PTSD induction by a trauma exposure on day 1 and a trigger exposure on day 2, to induce a PTSD-like phenotype in part of the mice, whereas others are resilient and do not display any PTSD-related symptoms. No difference in fear generalisation (assessed by freezing behaviour to a novel context) between PTSD-like and resilient mice was found prior to PTSD development. However, PTSD-like mice did display suppressed corticosterone responses to stress prior to PTSD development. In line with previous research reporting on abnormal hippocampal dysfunction in PTSD patients, we found that hippocampal blood flow, representing hippocampal activity, was significantly reduced in mice who had acquired a PTSD-like phenotype, which was not observed in resilient animals. Moreover, spine density in the ventral DG – critical for the process of pattern separation, preventing generalisation of memories – was significantly decreased in mice with a PTSD-like phenotype compared to resilient mice. Though we observed several indications for hippocampal dysfunction as pathology once a PTSD phenotype had been developed, none of the findings indicate hippocampal dysfunction as vulnerability factor before PTSD development (i.e., it is not a predisposition).

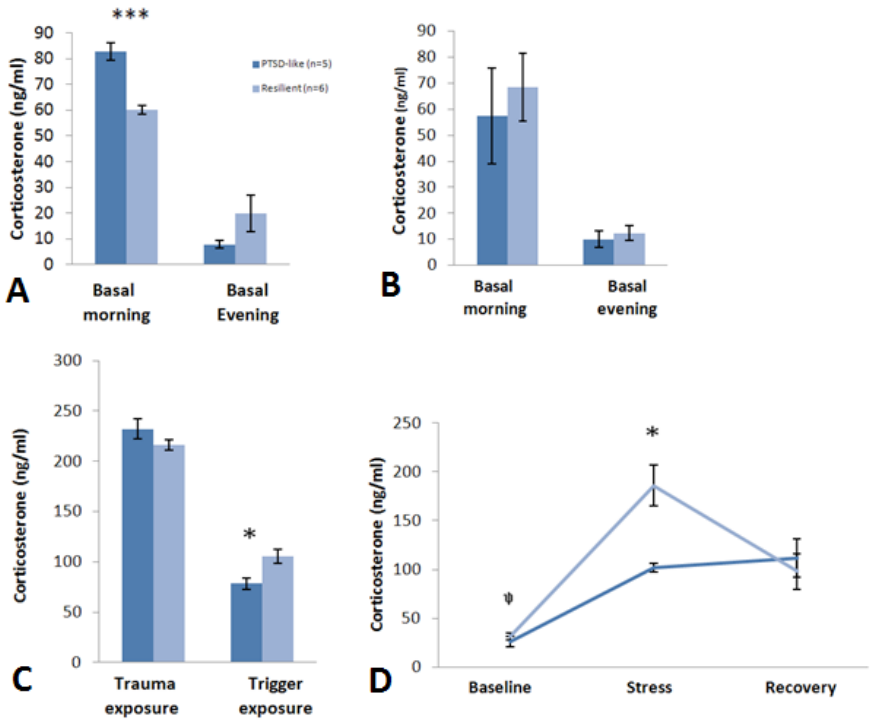

Fig. 6. Corticosterone levels in mice from experiment 1. Top graphs: Basal measurements before PTSD induction (**A**), and after PTSD induction (**B**). Bottom graphs: Stress-induced CORT levels (**C** and **D**). PTSD-like mice had an increased corticosterone level during pre-exposure morning ($p = .002$) measurement (D). Resilient mice had a significantly increased corticosterone level during trigger exposure ($p = .02$; C), during restraint stress ($p = .01$; D) and a trend-level significant increase during baseline measure of restraint stress ($p = .06$; D). *$p < .05$, ψ$p = .06$. Error bars indicate standard errors of the mean (*SEM*).

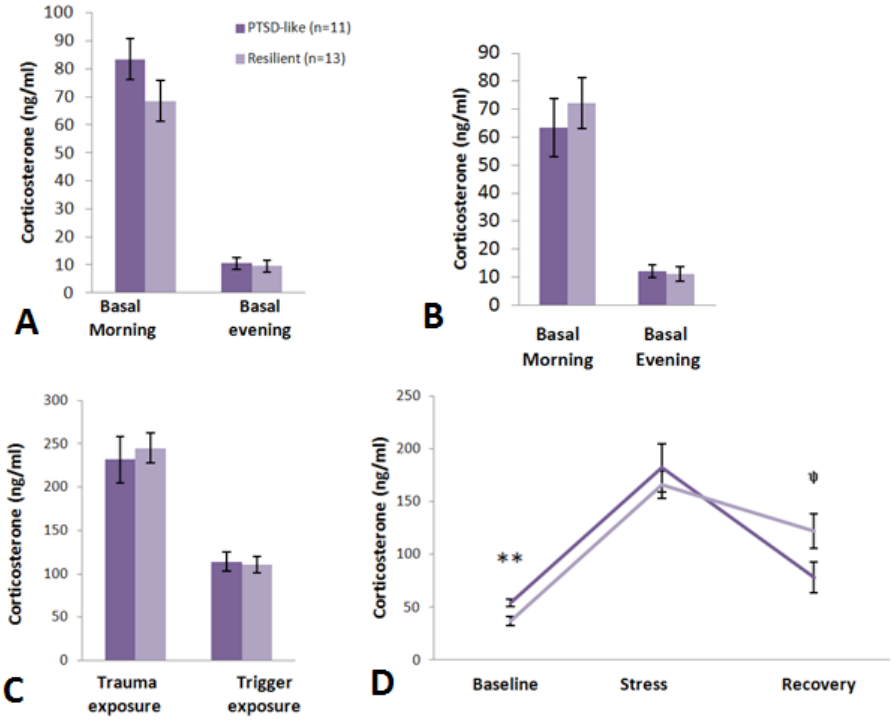

Fig. 7. Corticosterone levels from experimental group 2. Top graphs: Basal measurements before PTSD induction (**A**), and after PTSD induction (**B**). Bottom graphs: Stress-induced CORT levels (**C** and **D**). PTSD-like mice had a significantly increased baseline restraint stress corticosterone level (p = .006; D). However, resilient mice had a trend-level significant increase in corticosterone during restraint stress recovery (p = .06; D). **p < .01, ψp = .06. Error bars indicate *SEM*.

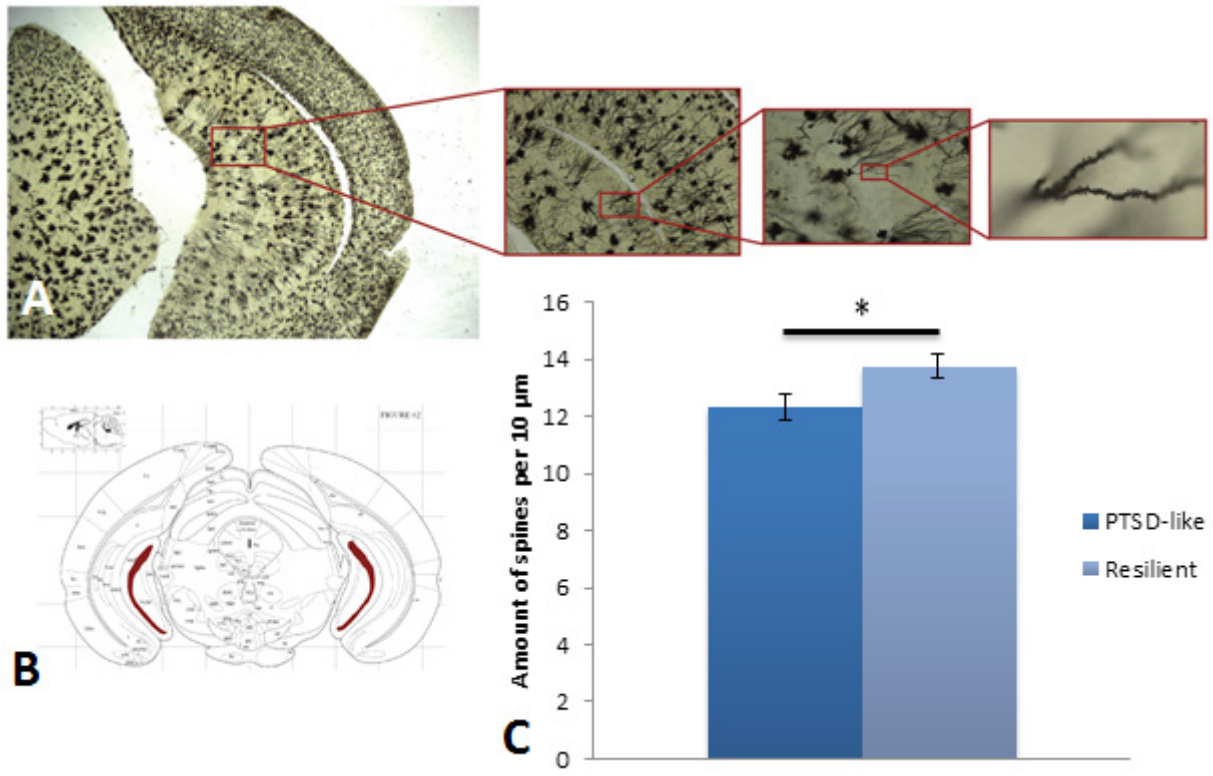

Fig. 8A. Golgi-Cox staining in brain tissue (from left to right: 5×, 20×, 40×, and 100× magnification). **B.** Mouse Brain Atlas drawing of ventral DG. Bregma coordinates = -3.28:-3.80 mm. **C.** Spine density per 10 µm in the ventral DG. Spine density was significantly reduced (p = .05) in PTSD-like mice compared to resilient mice. *p = .05. Error bars indicate *SEM*.

During the PTSD induction protocol, we exposed the animals to a trauma on the first day and a trigger on the next day. To measure fear generalisation, we measured the freezing response to the trigger context, which was an unfamiliar context to the animals, but contained a feature from the trauma context (i.e., the grid). We found that all mice of the experimental groups froze faster (during the first minute of trigger exposure, before the first shock) compared to a control group who underwent the same protocol but did not receive shocks during trauma exposure (*M* = 114.15, *SD* = 100.44; Henckens et al., unpublished data). This might indicate that the mice are in general more fearful in new environments (novelty-induced anxiety), or that the mice associate the trigger context with the trauma exposure context of the previous day. However, in this study, no reliable differences could be found between the PTSD-like and resilient animals. Therefore, we conclude that there is no indication of fear generalisation during PTSD induction in either PTSD-like or resilient animals. This is in line with research from Golub et al. (2005), in which they did not find an increase in freezing in a novel context containing a feature from the conditioning context (grid). They explain that cues most proximal to the trauma are often not directly associated with it (Ehler, Hackmann, & Michael, 2004; Golub et al., 2005).

Furthermore, in the first experimental group, we found a suppressed stress induced corticosterone level in PTSD-like animals, to restraint stress once the pathology had established. The findings of suppressed stress-induced corticosterone levels are in line with findings from previous animal work (Lebow et al., 2012), and patient studies (Yehuda, Giller, Southwick, Lowy, & Mason, 1991; Yehuda et al., 1996) on PTSD, where it is suggested that an enhanced negative feedback of the HPA-axis is causing these decreased cortisol (corticosterone) levels (Yehuda et al., 1991). Such enhanced negative feedback is also in line with the findings of lower corticosterone levels in PTSD-like animals during the recovery phase of restraint stress in group 2. Interestingly, we also found initial evidence for a suppressed corticosterone response to trigger exposure in PTSD-like mice compared to their resilient counterparts prior to PTSD development. Findings that PTSD-like mice display suppressed corticosterone responses already during PTSD induction seem to fit the findings of Zohar et al.

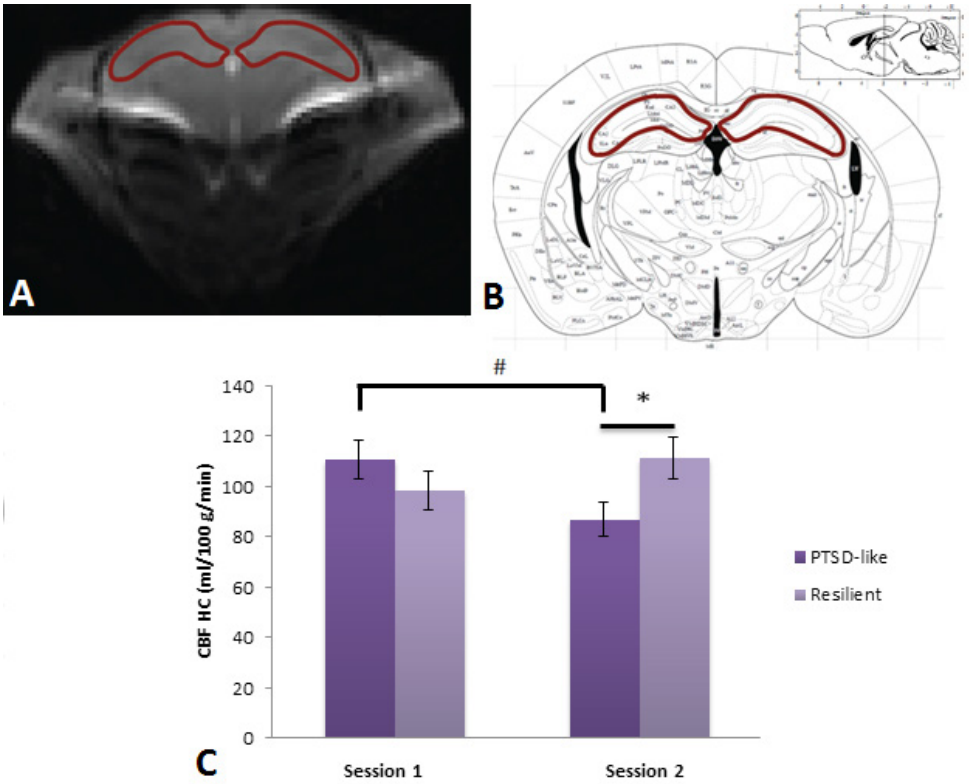

Fig. 9A. ASL MRI scan of the hippocampus, Bregma coordinate = -1.94 mm. **B.** Mouse Brain Atlas showing drawn region of interest (hippocampus), Bregma coordinate = -1.94 mm. **C.** Cerebral blood flow in the hippocampus in PTSD-like and resilient mice prior to and following PTSD induction. Blood flow was significantly reduced in PTSD-like mice after PTSD induction compared to resilient mice (p = .05). Furthermore, there was a trend-level significant time × group interaction in hippocampal CBF levels due to the decrease in CBF in PTSD-like only (p = .06). *p = .05 (t-test), #p = .06 (repeated-measures ANOVA). Error bars indicate *SEM*.

(2011), which suggested that the administration of a single high dose of corticosteroids immediately after trauma exposure can actually reduce symptoms of acute stress and post-traumatic stress. It was found that high-dose corticosterone in polytrauma patients administered immediately after trauma, reduced symptoms of PTSD, and in the animal study, high-dose corticosterone stimulated dendritic outgrowth and spine density in the DG (Zohar et al., 2011). Thus, immediately after trauma, corticosterone can counteract the symptomatological and morphological effects of PTSD. These findings were explained by a proposed role for high corticosterone levels immediately after trauma in recalibrating the HPA-axis in order to ensure appropriate regulation of the HPA-axis following trauma (Zohar et al., 2011). Also, corticosterone has been reported to facilitate the acquisition of contextual memory (Cordero, Venero, Kruyt, & Sandi, 2003). So, this indicates that corticosterone improves context-dependent memory and counteracts fear generalisation (Cordero et al., 2003; Pugh, Tremblay, Fleshner, & Rudy, 1997). This is also shown by Albrecht, Caliskan, Oitzl, Keinemann, & Stork (2013), who found that high levels of corticosterone infused in the ventral hippocampus following fear exposure have an anxiolytic effect following fear exposure (Albrecht et al., 2013). This might be in line with our result of increased corticosterone levels in resilient mice during trigger exposure. However, another seminal paper by Kaouane et al. (2012) reported that infusion of corticosterone in the hippocampus after fear conditioning actually induces PTSD-like memory impairments, which was reflected in a flawed memory for the correct predictor of the threat in mice (which thus generalised their fear).

Unfortunately, findings on suppressed stress-induced corticosterone levels were not replicated in the second experimental group, which could be the consequence of the differential experimental time line and repeated stress exposure in this group. Although maximal effort was put into reducing the stress induced by the anaesthesia required for MRI scanning, animals most likely still experienced moderate stress. Moreover, isoflurane, the anaesthesia required for MRI scanning, has been shown to induce a significant corticosterone response (Altholtz, Fowler, Badura, & Kovacs, 2006). Also, the restraint and basal corticosterone measures in this group were obtained at slightly different time points during the protocol. In sum, we observed a suppressed stress-induced corticosterone level in PTSD-like mice which could be the consequence and a predictor of an enhanced negative feedback of the HPA-axis in PTSD.

In addition to fear generalisation and neuroendocrine changes associated with PTSD development, we also assessed ventral DG neuronal morphology in PTSD-like versus resilient mice. The spine density analysis focused on the ventral DG, because it is associated with pattern separation and fear generalisation, and because it is involved in regulating affect, motivational, and emotional behaviour and the hormonal stress response (Fanselow & Dong, 2010; Kheirbek et al., 2013). We observed a significant decrease in ventral DG spine density in PTSD-like mice, suggesting a decreased excitatory input (Amaral et al., 2007) on the dendrites of the ventral DG granule cells in these animals compared to their resilient counterparts. These alterations in morphology might be a maladaptive response to trauma and can underlie the physiological and behavioural changes observed due to decreased excitatory input into the ventral DG. Previous research has produced some conflicting data on the effects of stress and trauma on spine density in the dorsal DG. On one hand, a rather acute increase in spines density in the dorsal DG has been found in response to acute stress (1 hour platform stress) after which the rats were immediately sacrificed (Sebastian, Estil, Chen, Schrott, & Serrano, 2013), while other studies have reported decreased dorsal DG spine density in response to stress on the longer time scale (8 days following single acute predator scent stress) in the DG (Cohen et al., 2014; Zohar et al., 2011). The current study is the first to describe the long-term effects of PTSD-like symptoms in the ventral DG.

The observed reduction in spine density could have several causes. Firstly, spine loss can be the result of a widespread loss of neurons in the amygdala and hippocampus (Fanselow & Dong, 2010; Kheirbek et al., 2012) and their axonal inputs on spines. However, compensatory increases in spine density in response to the loss of synaptic input have also been found (Fiala, Spacek, & Harris, 2006). Secondly, it might be caused by a local loss of neurogenesis or axonal sprouting (Fiala et al., 2006), known to be affected by stress (Altman & Das, 1967; McEwan, 1999; Redila & Christie, 2006). The exact reasons for spine density reduction in this study thus remain unclear. Further research must be conducted to investigate the origin of this loss of spines in the ventral DG in animals with a PTSD-like phenotype. Also, as these invasive measures of neuronal morphology could only be obtained after sacrificing the animals, it remains unknown whether the examined reductions in spine density in PTSD-

like mice are a cause or a consequence of PTSD development.

In PTSD patients, a relationship has been reported between reduced hippocampal activity and severity of PTSD symptoms (Astur et al., 2006). However, literature is conflicting on whether the reduced hippocampal activity is a predisposition or pathology (Hayes et al., 2011). Prospective studies where large groups of subjects (for example combat soldiers) are followed over time (Admon, Milad, & Hendler, 2013) provided initial evidence that reduced hippocampal activity in PTSD is a pathology rather than predisposition (Hayes et al., 2011). Here, using a mouse-model for PTSD, we show that differences in hippocampal blood flow as a measure of hippocampal activity (Zerbi et al., 2014), are not apparent before trauma induction, but become significant only after the PTSD induction protocol. This might indicate that neuronal alterations taking place following trauma and trigger exposure are responsible for this endophenotype. Potentially, this reduction of hippocampal blood flow following PTSD development might be caused by the physiological effects of the chronic stress experienced due to PTSD symptomatology, since chronic stress has been related to increased susceptibility to excitatory amino acids such as glutamate and the inhibition of growth factors like brain-derived neurotrophic factor (BDNF; Carrion, Haas, Garrett, Song, & Reiss, 2010).

Some limitations to this study should be mentioned as well. Findings on corticosterone levels in experiment 1 and experiment 2 were not replicated between the groups. As mentioned before, this might be the result of differences in the experimental design of the two groups. The restraint corticosterone and post-exposure basal corticosterone blood sample collections were obtained on different time points in both groups (they were reversed). For experiment 1, we strictly adhered to the previously validated timeline of PTSD induction, testing, and time of sacrifice (Lebow et al., 2012; Henckens et al., unpublished data). For experiment 2 we decided to deviate from this protocol, as we considered the second MRI session the main measurement of interest, which is why it should be performed immediately after the completion of the assessment of the behavioural phenotype (i.e., around the time of sacrifice in experiment 1). We did not want to delay this measurement, nor confound it by the additional exposure to restraint stress. The corticosterone stress response to restraint was considered of lesser importance to this experiment, which is why it was delayed. Also, additional stress caused by the anaesthesia used during the MRI scanning of group 2 may be responsible for the apparent inconsistencies. Further research is needed to replicate the association between stress-induced corticosterone levels and PTSD symptomatology, and to examine whether the differences found in the corticosterone results are in fact time-dependent (i.e., dependent on the specific moment of sampling). Another limitation is that we only investigated the density of spines (i.e., total amount of spines), but we did not investigate differences in the spine structure itself.

Changes in spine structure and morphology, for example spine size or shape alterations, might indicate reductions in functional integrity and synaptic activity (Fiala et al., 2006). Future research should examine whether there are apparent differences in spine structure and morphology after PTSD development. Furthermore, in this study, we only focused on spine density in the ventral DG, because the ventral DG is associated in pattern separation and regulation of affect, motivation and emotional behaviour. However, others also found changes in spine density in other areas of the hippocampus in the rodent brain in response to stress. An increase in spine density has been found in the hippocampal CA1 (Adamec et al., 2012; Diamond et al., 2006) and CA3 (Dias et al., 2014; Redila & Christie, 2006) after acute stress and in the amygdala in response to chronic stress (Astur et al., 2006), also a reduction in spine density was found in the medial prefrontal cortex in response to chronic stress (Hayes et al., 2011). Therefore, future research should assess whether the reduction in spine density found in the ventral DG in PTSD-like animals are also apparent in other hippocampal subregions (dorsal DG, CA3, etc.) or other PTSD-related brain regions. Lastly, we only assessed blood flow through the complete hippocampus, without separating the several subnuclei. However, the hippocampus consists of several subnuclei with different functions (Amaral & Lavenex, 2006), and others found reduced hippocampal activity in the posterior and, to a lesser extend in the anterior, hippocampus in response to traumatic pictures in PTSD patients (Hayes et al., 2011). Therefore, the specificity of the findings remains to be elucidated. Also, blood flow in subregions of the hippocampus, and especially the ventral and dorsal DG should be examined in future studies. As it has been found that in both humans and mice, sex differences exist in stress susceptibility and anxiety-like behaviour after trauma exposure, where females seem to be more susceptible to develop stress-related disorders

(Adamec, Head, Blundell, Burton, & Berton, 2006; Shansky, 2015).

Here, we chose to study the population (i.e., males) displaying the most robust and stable response to physical stressors as a proof of principle (Ter Horst et al., 2012). Future studies should assess whether similar effects in terms of functional and morphological changes in the hippocampus as a consequence of PTSD are present in the female brain as well. Lastly, we studied inter-individual differences in PTSD vulnerability in animals with a nearly identical genetic background. Although clear inter-individual differences were observed, previous reports have indicated that gene-environment interactions render certain individuals more vulnerable to the development of PTSD after trauma exposure than others (Mehta & Binder, 2012), a factor currently neglected in this study. Future assessment of such gene-trauma interactions is of critical importance to enhance our insight in the underlying mechanisms of aberrant hippocampal function.

Conclusion

Here, we investigated the association between hippocampal function and associated fear generalisation in PTSD, to find out when exactly in disease development aberrant hippocampal functioning emerges. We found no indication of differential fear generalisation during the PTSD induction protocol predicting the subsequent development of PTSD-like symptomatology. The only index predicting PTSD development was the reduced corticosterone response to trigger exposure. These data suggest that corticosterone secretion during trauma exposure might be a compensatory mechanism (which is operational in resilient animals) to protect against the harmful effects of trauma and trigger exposure. This protective effect of increased corticosterone after trauma might be by recalibration of the HPA-axis in order to ensure appropriate regulation of the HPA-axis and proper memory for the trauma context, and thus, counteracting fear generalisation. In line with literature, we observed reduced hippocampal blood flow in PTSD-like compared to resilient animals once pathology had been established. However, this difference only became apparent after PTSD development, indicating the reflection of pathology rather than a predisposition. Moreover, spine density in the ventral DG was reduced in PTSD-like mice. This might indicate reduced excitatory input due to PTSD development, which might cause impairments in pattern separation and ultimately cause the phenotype of fear generalisation in PTSD. Although further research is needed to obtain more insight in the exact mechanisms underlying the observed effects, these findings indicate that the hippocampus plays a significant role in the development of PTSD and aberrant fear generalisation, and therefore significantly contribute to the symptomatology of PTSD.

Acknowledgments

The completing of this research paper could not have been made possible without the contribution of several people to whom I would like to express my gratitude. Firstly, I would like to thank Dr. Marloes Henckens for her supervision and support during this research project. Furthermore, I would like to thank Gil Schulte and Miranda van Bodegom for their experimental work during MRI acquisition, and Prof. Dr. Tamas Kozicz as the second reader of the thesis. Lastly, I would like to thank Dr. Judith Homberg for the opportunity to do my MSc research internship at her department.

References

Adamczak, J., Farr, T., Seehafer, J., Kalthoff, D., & Hoehn, M. (2010). High field BOLD response to forepaw stiumulation in the mouse. *NeuroImage*, *51*, 704-712.

Adamec, R., Head, D., Blundell, J., Burton, P., & Berton, O. (2006). Lasting anxiogenic effects of feline predator stress in mice: Sex differences in vulnerability to stress and predicting severity of anxiogenic response from the stress experience. *Physiology and Behaviour, 88*(1-2), 12-29.

Adamec, R., Hebert, M., Blundell, J., & Mervis, R. (2012). Dendritic morphology of amygdala and hippocampal neurons in more and less predator stress responsive rats and more and less spontaneously anxious handled controls. *Behavioural Brain Research, 226*(1), 133-146.

Admon, R., Milad, M., & Hendler, T. (2013). A causal model of post-traumatic stress disorder: disentangling predisposition from acquired neural abnormalities. *Trends in Cognitive Sciences, 17*(7), 337-347.

Aimone, J., Deng, W., & Gage, F. (2011). Resolving new memories: a critical look at the dentate gyrus, adult neurogenesis and pattern separation. *Neuron, 70*(4), 589-596.

Albrecht, A., Caliskan, G., Oitzl, M., Keinemann, U., & Stork, O. (2013). Long-lasting increase of corticoterone after fear memory reactivation: anxiolytic effects and network activity modulation in the ventral hippocampus. *Neuropsychopharmocology, 38*, 386-394.

Altholtz, L., Fowler, K., Badura, L., & Kovacs, M. (2006). Comparison of the stress response in rats to repeated isoflurane or CO2:O2 anesthesia used for restraint during serial blood collection via jugular vein. *Journal of the American Association for Laboratory Animal Science, 45*(3), 17-22.

Altman, J., & Das, G. (1967). Postnata neurogenesis in the guinea-pig. *Nature, 214*, 1098-1101.

Amaral, D., & Lavenex, P. (2006). Hippocampal neuroanatomy. In D. Amaral, & P. Lavenex (Eds.), *The Hippocampus Book*. Oxford, England: Oxford University Press.

Amaral, D., Scharfman, H., & Lavenex, P. (2007). The dentate gyrus: Fundamental neuroanatomical organization (dentate gyrus for dummies). *Progress in Brain Research, 163,* 3-22.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Press.

Astur, R., St.Germanin, A., Tolin, D., Ford, J., Russell, D., & Stevens, M. (2006). Hippocampus function predicts severity of post traumatic stress disorder. *CyberPsychology and Behaviour, 9*(2), 234-240.

Bohlen, V., & Halbach, O. (2009). Structure and function of dendritic spines within the hippocampus. *Annals of Anatomy, 191*, 518-531.

Bonne, O., Brandes, D., Gilboa, A., Mosche Gomori, J., Shenton, M., Pitman, R., & Shalev, A. (2001). Longitudinal MRI study of hippocampal volume in trauma survivors with PTSD. *American Journal of Psychiatry, 158*, 1248-1251.

Bonne, O., Vythilingam, M., Ingaki, M., Wood, S., Neumeister, A., Nugent, A., ... Chamey, D. (2008). Reduced posterior hippocampal volume in posttraumatic stress disorder. *Journal of Clinical Psychiatry, 69*(7), 1087-1091.

Breslau, N. (2001). The epidemiology of posttraumatic stress disorder: What is the extent of the problem? *Journal of Clinical Psychiatry, 62*, 16-22.

Carrion, V., Haas, B., Garrett, A., Song, S., & Reiss, A. (2010). Reduced hippocampal activity in youth with posttraumatic stress symptoms: an fMRI study. *Journal of Pediatric Psychology, 35*(5), 559-569.

Cohen, H., Kozlovsky, N., Matar, M., Zohar, & Kaplan, Z. (2014). Distinctive hippocampal and amygdalar cytoarchitectural changes underlie specific patterns of behavioural disruption following stress exposure in an animal model of PTSD. *European Neuropsychopharmacology, 24*, 1925-1944.

Cordero, M., Venero, C., Kruyt, N., & Sandi, C. (2003). Prior exposure to a single stress session facilitates subsequent contextual fear conditioning in rats, evidence for a role of corticosterone. *Hormones and Behaviour, 44*, 338-345.

Diamond, D., Campbell, A., Park, C., Woodson, J., Conrad, C., Bachstetter, A., & Mervis, R. (2006). Influence of predator stress on the consolidation versus retrieval of long term spatial memory and hippocampal spinogenesis. *Hippocampus, 16*, 571-576.

Dias, G., Bevilaqua, M., Da Luz, A., Fleming, R., Carvalho, L., Cocks, G., ... Gardina, P. (2014). Hippocampal biomarkers of fear memory in an animal modal of generalized anxiety disorder. *Behavioural Brain Research, 263,* 34-45.

Ehler, A., Hackmann, A., & Michael, T. (2004). Intrusive re-experiencing in post traumatic stress disorder: Phenomenology, theory and therapy. *Memory, 12*(4), 403-415.

Fanselow, M., & Dong, H. (2010). Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron, 65*, 7-19.

Fiala, J., Spacek, J., & Harris, K. (2006). Dendritic structure and complexity in adult hippocampal dentate gyrus. *Neuroscience, 137*, 1299-1307.

Golub, Y., Mauch, C., Dahlhoff, M., & Wotjak, C. (2005). Consequences of extinction training on associative and non-associative fear in a mouse model of posttraumatic stress disorder (PTSD). B*ehavioural Brain Research, 205*, 544-549.

Grandjean, J., Schroeter, A., Batata, I., & Rudin, M. (2014). Optimization of antesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. *NeuroImage, 102*, 838-847.

Greenberg, T., Carlson, J., Cha, J., Hajcak, G., & Mujica-Parodi, L. (2013). Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. *Depression and Anxiety, 30*, 242-250.

Hayes, J., Labar, K., McCarthy, G., Selgrade, E., Nasser, J., Dolcos, F., & Morey, R. (2011). Reduced hippocampal and amygdalar activity predicts memory distortions for trauma reminders in combat-related PTSD. *Journal of Psychiatric Research, 45*, 660-669.

Ji, J., & Maren, S. (2008). Differential roles for hippocampal areas CA1 and CA3 in the contextual encoding and retrieval of extinguished fear. *Learning & Memory, 15*, 244-251.

Kaouane, N., Porte, Y., Valle, M., Brayda-Bruno, L., Mons, N., Calandreau, L., ... Desmedt, A. (2012). Glucocorticoids can induce PTSD-like memory impairments in mice. *Science, 335*, 1510-1513.

Kavushansky, A., Vouimba, R., Cohen, H., & Richter-Levin, G. (2006). Activity and plasticity in CA1, the dentate gyrus and the amygdala following controllable vs. uncontrollable water stress. *Hippocampus, 16*, 35-42.

Kessler, R., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. (1995). Posttraumatic Stress Disorder in the National Comorbidity Survey. *Archives of General Psychiatry Journal, 52*, 1048-1060.

Kheirbek, M., Drew, L., Burghardt, N., Costantini, D., Tannenholz, L., Ahmari, S., ... Hen, R. (2013). Differential control of learning and anxiety along the dorsoventral axis of the dentate gyrus. *Neuron, 77,* 955-968.

Kheirbek, M., Klemenhagen, K., Sahay, A., & Hen, R. (2012). Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. *Nature*

Neuroscience, 15, 1613-1620.

Lebow, M., Neufeld-Cohen, A., Kuperman, Y., Tsoory, M., Gil, S., & Chen, A. (2012). Susceptibility to PTSD-like behaviour is mediated by corticotropin-releasing factor receptor type 2 levels in the bed nucleus of stria terminalis. *The Journal of Neuroscience, 32*(20), 6906-6916.

McEwan, B. (1999). Stress and hippocampal plasticity. *Annual Review of Neuroscience, 22*, 105-122.

Mehta, D., & Binder, E. (2012). Gene x environment vulnerability factors for PTSD: The HPA-axis. *Neuropharmocology, 62,* 654-662.

Nimchinsky, E., Sabatini, B., & Svoboda, K. (2002). Structure and function of dendritic spines. *Annual Review of Physiology, 64*, 313-353.

Pitman, R., Rasmusson, A., Koenen, K., Shin, L., Orr, S., Gilbertson, M., ... Liberzon, I. (2012). Biological Studies of Post-traumatic Stress Disorder. *Nature Reviews Neuroscience, 13,* 769-787.

Pugh, R., Tremblay, D., Fleshner, M., & Rudy, J. (1997). A selective role for corticosterone in contextual-fear conditioning. *Behavioural Neuroscience, 111*(3), 503-511.

Redila, V., & Christie, B. (2006). Exercise-induced changes in dendritic structure and complexity in adult hippocampal dentate gyrus. *Neuroscience, 137*, 1299-1307.

Sahay, A., Wilson, D., & Hen, R. (2011). Pattern separation: A common function for new neurons in the hippocampus and olfactory bulb. *Neuron, 70*, 582-588.

Sebastian, V., Estil, J., Chen, D., Schrott, L., & Serrano, P. (2013). Acute physiological stress promotes clustering of synaptic markers and alters spine morphology in the hippocampus. *Plos One, 8*(10), e79077.

Servatius, R., Ottenweller, J., & Natelson, B. (1995). Delayed startle sensitization distinguishes rats exposed to one or three stress sessions: Further evidence towards an animal model of PTSD. *Biological Psychiatry, 38*, 589-546.

Shansky, R. (2015). Sex differences in PTSD resilience and susceptibility: Challenges for animals models of fear learning. *Neurobiology of Stress, 1*, 60-65.

Shin, L., Rauch, S., & Pitman, R. (2006). Amygdala, medial prefrontal cortex and hippocampul function in PTSD. *Annals of New York Academy of Sciences, 1071*, 67-79.

Shin, L., Shin, P., Heckers, S., Krangel, Y., Macklin, M., Orr, S., ... Rauch, S. (2004). Hippocampal function in posttraumatic stress disorder. *Hippocampus, 14*, 292-300.

Shors, T. J., & Leuner, B. (2004). The opposite effects of stress on dendritic spines in male vs. female rats are NMDA receptor-dependant. *European Journal of Neuroscience, 19*(1), 145-150.

Ter Horst, J., De Kloet, E., Schachinger, H., & Oitzl, M. (2012). Relevance of stress and female sex hormones for emotion and cognition. *Cellular and Molecular Neurobiology, 32*, 725-735.

Van Zuiden, M., Geuze, E., Willemen, H., Vermetten, E., Maas, M., Amarouchi, K., ... Heijnen, C. (2012). Glucocorticoid receptor pathway components predict posttraumatic stress disorder symptom development: a prospective study. *Biological Psychiatry, 71*, 309-316.

Yehuda, R., Giller, E., Southwick, S., Lowy, M., & Mason, J. (1991). Hypothalamic-pituitary-adrenal dysfunction in posttraumatic stress disorder. *Biological Psychiatry, 30*, 1031-1048.

Yehuda, R., Teicher, M., Trestman, R., Levengood, R., & Siever, L. (1996). Cortisol regulation in posttraumatic stress disorder and major depression: a chronobiological analysis. *Biological Psychiatry, 40,* 79-88.

Zerbi, V., Wiesman, M., Emmerzaal, T., Jansen, D., Van Beek, M., Mutsaers, M., ... Kiliaan, A. (2014). Resting-state functional connectivity changes in aging apoE4 and apoE-KO mice. *Neurobiology of Disease, 34*(42), 13963-13975.

Zohar, J., Yahalom, H., Kozlovsky, N., Cwikel-Hamzany, S., Matar, M., Kaplan, Z., ... Cohen, H. (2011). High dose hydrocortisone immediately after trauma may alter the trajectory of PTSD: Interplay between clinical and animal studies. *European Neuropsychopharmacology, 21*, 796-809.

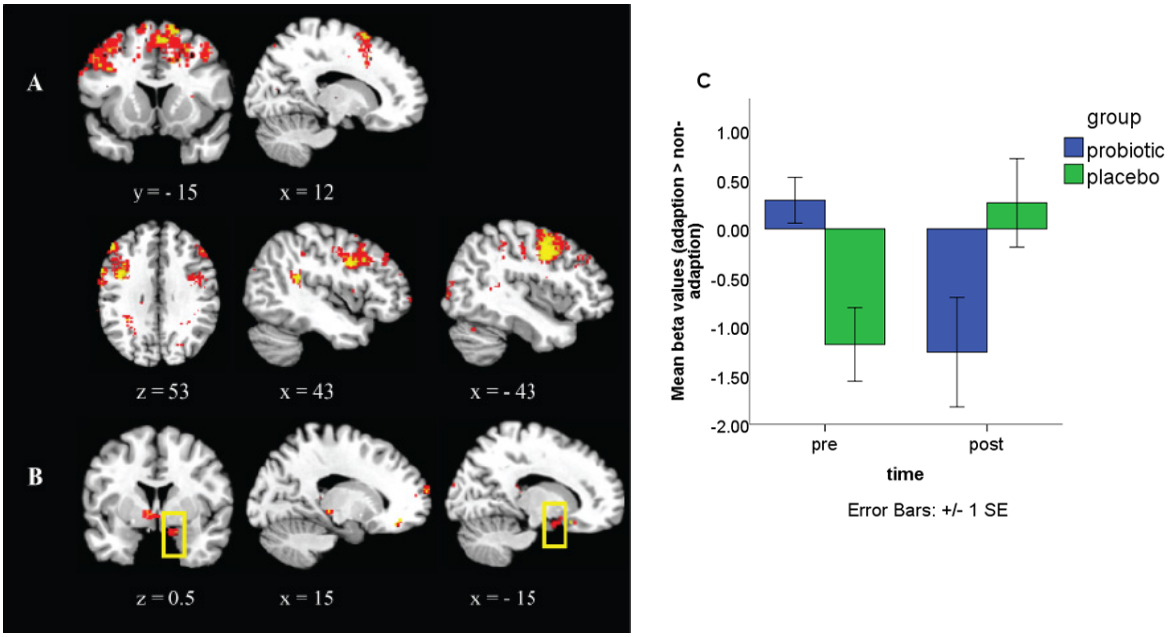

Fig. 5. Emotional face Stroop: positive task effects and effects of probiotic intervention on brain activity. **A.*** Positive task effects of emotional Stroop adaptation (adaptation > non-adaptation). Here, brain regions are displayed showing more activity during adaptation compared with non-adaptation trials. **B.*** Effects of probiotic intervention on the brain during emotional Stroop adaptation (adaptation > non-adaptation) (Probiotic > Placebo, Pre > Post). Here, brain regions are displayed showing more activity for the probiotic relative to the placebo group in the pre- compared with the post-session.
C. Extracted mean beta values (adaptation > non-adaptation) from an ROI created for the amygdala (MNI coordinates: -16 4 -22) based on the functional images received from the Group × Time interaction contrast on (Probiotic > Placebo, Pre > Post) at $p < .001$ (uncorrected).
* Results are displayed at exploratory thresholds in *yellow* at $p < .001$ (uncorrected) and in *red* at $p < .005$ (uncorrected). Images are shown in radiological orientation, left = right (MNI coordinates).

regions in the frontal cortex including vmPFC were shown to be *deactivated* for the probiotic versus placebo group after the intervention, contrary to our hypothesis. Minor amygdala *deactivation* could be observed as well ($p < .001$, uncorrected; more clearly at $p < .005$, uncorrected) when groups were compared after intervention, being in line with our hypothesis (see Fig. 5B). For illustrative purposes we created a region of interest (ROI) for this region of activation (see Fig. 5C) from functional images received from the Group × Time interaction contrast (Probiotics > Placebo, Pre > Post; $p < .001$, uncorrected). Mean beta values (adaptation > non-adaptation) were extracted showing a *decrease* in amygdala activity for the probiotic group (Pre > Post), as expected. The opposite could be observed for the placebo group.

Classic colour-word Stroop paradigm. Main task effects of the Stroop effect revealed significant activations for regions in left lateralised frontal regions (see Table 3 and Fig. 6A). We did not observe significant effects of the probiotic intervention on brain regions during this paradigm, however at an exploratory threshold of $p < .005$ (uncorrected), activation of lPFC (i.e., IFG) for the probiotic group compared with the placebo group (Post > Pre) could be observed in line with our hypothesis. Additionally, we found vmPFC activity for this contrast (see Fig. 6B). For illustrative purposes we created an ROI for the area in the lPFC (see Fig. 6C) based on the functional images received from the Group × Time interaction contrast (Probiotic > Placebo, Post > Pre; $p < .005$, uncorrected). Mean beta values were extracted showing an increase in activity for the probiotic group, as expected (Post > Pre). The placebo group showed effects in the opposite direction.

Discussion

In the present study our aim was to investigate the effects of a multispecies probiotic on brain functioning and its underlying mechanisms. We were particularly interested in whether the probiotic effects would extend beyond emotion processing and associated brain regions, also affecting brain regions involved in emotion regulation and general cognitive control processes.

Table 2.

Main task effects of emotional Stroop adaptation (adaptation – non-adaptation) and (non-adaptation – adaptation). Clusters showing greater BOLD* activity for adaptation trials compared to non-adaptation trials (whole brain corrected $p_{FWE} < .05$). MNI stereotactic coordinates of local BOLD maxima.

Region	cluster *p* (FWE-corrected)	cluster equivk	peak T	x, y, z {mm}
Emotional Stroop adaptation				
(adaption > non-adaptation)				
Left precentral gyrus	.000	441	6.24	-38, -2, 44
Left precentral gyrus			5.96	-46, -2, 38
Left precentral gyrus			5.25	-38, 0, 36
Right precentral gyrus	.001	189	5.30	42, 0, 32
Right inferior frontal gyrus pars opercularis			5.28	38, 16, 32
Right inferior frontal gyrus pars opercularis			4.31	46, 14, 34
Right pre-supplementary motor area	.002	167	4.93	6, 4, 68
Left pre-supplementary motor area			4.47	0, 8, 54
Left pre-supplementary motor area			4.36	-8, 16, 54
(non-adaptation > adaptation)				
No significant clusters				

*BOLD = Blood oxygenated level dependent

Main effects of tasks

Despite the small sample size studied here, we found sufficient task-related brain activations in line with findings of earlier studies, thus we can confidently conclude that the tasks used in our design functioned as expected. This is not very surprising since we used three robust paradigms in order to ensure that we could measure the effects we were interested in. The significant activation of thalamus during the emotional face matching task for instance, a region known to influence amygdala activity, is in line with findings by Hariri et al. (2000). Additionally, activation in bilateral inferior occipital gyri (IOG), a region shown to be involved in face processing, also indicated that this task worked well since it was found when brain activation for matching of emotional faces was compared with matching of shapes (Haxby, Hoffman, & Gobbini, 2000). Significant task-related brain activations in the emotional face Stroop task in medial and later PFC, particularly right IFG and bilateral pre-SMA, as well as *deactivation* of the vmPFC when adaptation trials were compared with non-adaptation trials were also in line with earlier studies using similar paradigms (Roberts & Hall, 2008; Etkin et al., 2006). Finally, the classic colour-word Stroop task activated medial and lateral frontal areas as well, here also including left pre-SMA and left IFG, regions often found to be activated during (Stroop) conflict paradigms (e.g., Roberts & Hall, 2008; Zoccatelli, Beltramello, Alessandrini, Pizzini, & Tassinari, 2010).

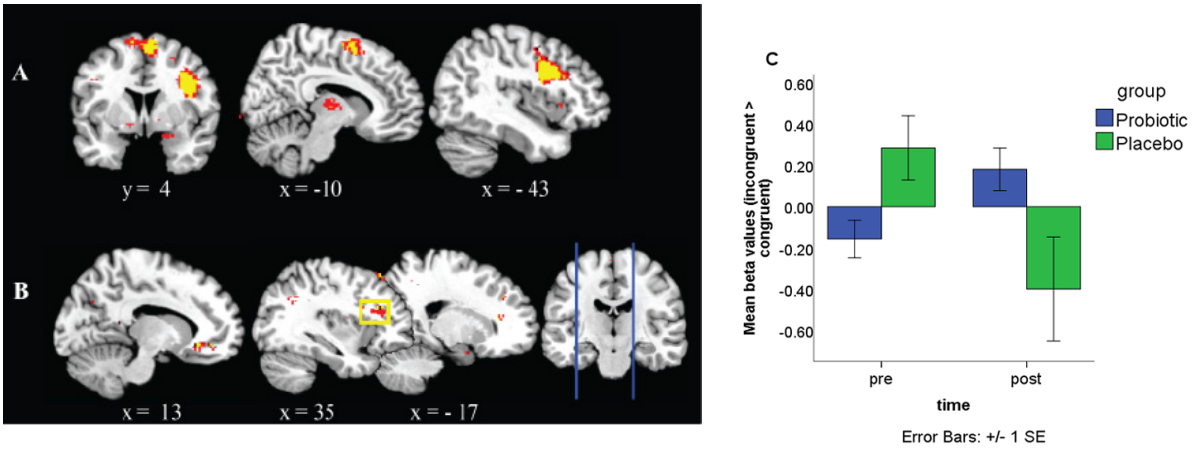

Fig. 6. Classic colour-word Stroop: positive task effects and effects of probiotic intervention on brain activity. **A.*** Positive task effects of Stroop effect (incongruent > congruent). Here, brain regions are displayed showing more activation during incongruent versus congruent trials. **B.*** Effects of probiotic intervention on the brain during Stroop effect (incongruent > congruent) (Probiotic > Placebo, Post > Pre). Here, brain regions are displayed showing more activation for the probiotic relative to the placebo group in the post- compared with the pre-session.
C. Extracted mean beta values (incongruent > congruent) from an ROI created for a region in the LPFC (MNI coordinates: 34 26 20) based on the functional images received from the Group × Time interaction contrast on (Probiotic > Placebo, Post > Pre) at $p < .005$ (uncorrected).
* Results are displayed at exploratory thresholds in *yellow* at $p < .001$ (uncorrected) and in *red* at $p < .005$ (uncorrected). Images are shown in radiological orientation, left = right (MNI coordinates).

Probiotic intervention effects

The probiotic intervention did not show significant effects on brain regions during the utilised paradigms, which is not surprising considering the very small sample size of $n = 6$ in the probiotic group and $n = 5$ in the placebo group. Yet, with exploratory, uncorrected thresholds indications of probiotic effects on brain activation in all three tasks could be seen of which some were in line with our hypotheses. Analyses showed an intervention-induced decrease in amygdala activity for the probiotic group compared with the control group only for the emotional face Stroop paradigm. Earlier studies have found the amygdala to be crucial for experiencing emotional conflict (Etkin et al., 2006). This preliminary finding could thus indicate that participants in the probiotic group experienced less emotional conflict in this task after the intervention compared with the placebo group. Additionally, this indication of *decreased* amygdala activity fits with what Tillisch et al. (2013) described in their results. However, they found this *deactivation* only in comparison with a non-intervention group than in comparison between the experimental group and the control group. This result might be due to the small group sizes investigated in their study or could indicate effects of the fermented milk product (here delivery vehicle) itself. In case our preliminary finding reaches significance with the final sample in our study, we could provide evidence for effects of probiotics on this affective brain region, unlikely induced by the delivery vehicle of the probiotics since here the probiotic and placebo group were compared.

In addition to decreased activity in the amygdala we expected to find enhanced activations in lPFC and mPFC areas during both Stroop paradigms due to the probiotic intervention. However, intervention-induced enhancements in lateral frontal activation, here left IFG, could be observed only for the classic colour-word Stroop paradigm. The IFG has been shown to be involved in cognitive control processes and to be important for inhibitory control (e.g., Swick, Ashley, & Turken, 2008). This preliminary finding could thus indicate an improvement of inhibitory control in the probiotic group by means of the probiotic intervention. Intervention-induced enhancement of vmPFC activation was also found in this paradigm for the probiotic group compared with the placebo group after the intervention. This finding is not in line with our hypotheses since this region has been shown to be more involved in cognitive control processes involving emotion whereas this paradigm does not include emotion stimuli (e.g., Winecoff et al., 2013). Nevertheless, in addition to many studies indicating its importance with respect to decisions involving

emotions or varying degrees of (un)certainty, it has also been indicated that this region is crucial for general decision making (Fellows & Farah, 2007). Therefore, these preliminary findings could hint towards an improvement of general decision making, for instance fewer impulsive decisions due to the probiotic-intervention. Intervention-induced *deactivations*, rather than the expected enhancement of vmPFC, were found in the emotional face Stroop paradigm as well as the emotional face matching paradigm; for the latter we did not predict any probiotic effects on prefrontal regions. Even though these preliminary results are not in line with our hypotheses, all regions found here were close to or overlapping with the regions activated during main task effects. This overlap indicates an intervention-induced *decrease* of activity of task-related regions during emotion processing and regulation.

Nevertheless, these preliminary brain-related findings, whether unexpected or in line with our hypothesis, need to be interpreted with caution due to the small sample size studied here, which is underlined by the findings of ROI beta value extraction. The mean beta values extracted from the ROIs created from the classic colour-word Stroop (lPFC) and the emotional face Stroop paradigm (amygdala) showed effects in line with our hypothesis for the probiotic group for both brain regions. Differences in activation in these regions could be seen when the baseline activation was compared to the activation at the post-session for the two groups, revealing group differences. However, group differences at baseline could also be observed. It is thus difficult to interpret these preliminary findings. With the final sample, unbiased statistical ROI analyses wherein ROIs are created based on anatomical images, will be performed in order to see by which group and time point the effects are driven.

Underlying gut-brain mechanisms

As we had a small sample size of eleven participants and therefore examined at exploratory, uncorrected thresholds, it is difficult to draw significant conclusions about underlying neurocognitive mechanisms in this study. With the complete sample of participants we hope to find effects of the probiotic on brain regions extending beyond emotion regions, which have already been indicated with this small sample size. In case these findings reach significance in the final sample it would be plausible to suggest a general gut-brain mechanism to underlie the effects such as immunological mechanisms or the metabolic pathway (described below). The immunological mechanisms (as described in the introduction) may lead to rather whole brain than very specific effects since this mechanism is involved in various processes in the body. Additionally, gut-bacteria can produce metabolites (metabolic pathway) that can enter the bloodstream thereby affecting the local enteric nervous system as well as the central nervous system, which also suggests whole brain effects. Interactions of neurochemicals between the central nervous system and gut microbiota have been shown to be bidirectional (e.g., Lyte, 2014). The produced metabolites can be precursors of neurotransmitters or can affect those travelling through the bloodstream and cross the blood-brain barrier to affect neurotransmitter synthesis in the brain (Collins, Surette, & Bercik, 2012). Some bacterial strains have been indicated to produce, for instance gamma-Aminobutyric acid (GABA) or tryptophan, a precursor of serotonin (O'Mahony, Clarke, Borre, Dinan, & Cryan, 2015). Tryptophan, for example, can affect various brain regions as serotonergic projections can be found throughout the brain, indicating that in case this mechanism is one to play a role here, effects would concern whole brain effects rather than simply affecting specific brain regions (e.g., Charnay & Leger, 2010).

Additionally, gut microbiota can produce hormones and regulate their secretion (Neuman, Debelius, Knight, & Koren, 2015), which is important concerning stress regulation, which has been shown to be malfunctioning in germ-free mice (Sudo et al., 2004). The immunological mechanism is tightly coupled with other mechanisms and systems in the body such as the metabolic mechanism and endocrine system. Through the regulation of hormone levels the gut-bacteria can, for instance affect the immune system (Neuman et al., 2015). In order to find out which one of these mechanisms plays a significant role in the current study, or whether they interact, it is necessary to include blood samples of participants in future studies. These samples can be used to measure the amount of inflammation markers in the blood of subjects providing a more direct measurement of the mechanism via which the probiotic changes in gut microbiome might affect cognitive functions. In case results concerning the PFC disappear with the final sample and effects are specific to affective brain regions, we might have reason to conclude another gut-brain mechanism to underlie the findings. In a study by Cryan and O'Mahony (2011), anxiolytic and antidepressant effects of a bacterium in mice were prevented when animals underwent vagotomy, suggesting a role

of the vagus nerve in gut-brain communication. The vagus nerve transmits signals from the enteric nervous system to the central nervous system. Its efferent pathway is crucial for the regulation of a number of cytokines in response to stress signals in the gut, suggesting a certain degree of overlap between the vagus nerve and the immune system (Sherman, Zaghouani, & Niklas, 2015). The vagus nerve sends signals about sensations occurring in the gut back to the brain providing the nucleus tractus sollitarius with gut-related information. Subsequently the information is transmitted to the parabrachial nucleus, which is connected to various brain regions (including insular, hypothalamus and amygdala) (King, 2007; Mayer, 2011) amongst others involved in emotion processing and regulation and maintenance of bodily homeostasis. Thus, in case the vagus nerve may play a major role in gut-brain communication here, we would expect more specific, direct effects on brain areas involved in emotion processing and regulation (rather than whole brain effects, including the PFC). Nevertheless, it is likely that these different mechanisms work together to a certain degree rather than functioning completely independently.

Limitations, strengths, and future directions

For future studies, it might be interesting to include a vagus nerve intervention. Current research is intended to test effects of blocking the vagus nerve in patients with obesity (e.g., Shikora et al., 2015). In a future study the vagus nerve could be blocked for a specific period of time for one group of participants, thereby adding a direct measurement of vagal nerve contribution to gut-brain communication. However, so far there are only invasive techniques available to block or stimulate the vagus nerve, which is not ethically appropriate for the present study. However, it is also possible to make predictions about vagal activity by means of a person's heart rate variability (HRV), more specifically the respiratory sinus arrhythmia (RSA) – a measurement of heart rate change in response to inhalation and exhalation – which was first proposed to be associated with vagal activity by Hering (1910) (as stated by Berntson et al., 1997). If the vagus nerve plays a role in gut-brain communication, changes in vagal tone should be observable (Alcock, Maley, & Aktipis, 2014).

In this preliminary data set one of the biggest limitations is the small sample size per group, which limits the amount of meaningful analyses that can be applied and decreases the power of related results. The final study will include 60 participants, 30 per group, which will lead to more reliable results and presumably less noise in the data. Additionally, due to time limits in the present study trials in the classic colour-word Stroop paradigm were evenly split into congruent and incongruent ones. However, Stroop interference has been shown to be stronger when overall proportion congruency is higher, that is, when the paradigm consists of proportionately more congruent than incongruent trials, which was seen in several studies (e.g., Kane & Engle, 2003; Logan & Zbrodoff, 1979). In future studies, additional congruent trials could thus be added in order to increase the Stroop interference and related brain activations. Furthermore, we aim at performing additional analyses with the complete sample such as brain-behaviour correlation analyses, in order to receive a more detailed picture of the data. In order to investigate whether the probiotic intervention affected affective brain regions directly or rather indirectly via the PFC, connectivity analyses between mPFC regions and the amygdala will be carried out for this paradigm with the complete sample of 60 participants.

A future strength of this study concerns the collected feces samples from participants. With the final sample we will apply analyses to the feces, enabling us to look for different bacteria strains in the sample. By means of a database it is then possible to find out which metabolites these bacteria can produce.

Conclusion

Taken together, the present study investigated the effects of a multispecies probiotic in a first, small subsample of participants. We can conclude that our tasks functioned as expected, although they did not show complete congruency with findings of earlier studies. Additionally, we could observe that indications of probiotic effects extended beyond affective brain regions as was expected. These effects included *decreased* vmPFC activity during emotion regulation, and more importantly, lPFC activity enhancement during a 'pure' non-emotional cognitive control task.

Clinical relevance

This study is of high clinical relevance since it aimed at finding effects of a probiotic on brain regions such as the amygdala involved in a variety

of affective psychiatric disorders (e.g., Peluso et al., 2009). Our preliminary results indicate a reduction of amygdala activity due to the probiotics in one of the tasks, which is of specific relevance for depression as depression disorder is often associated with hyperactivation of the amygdala (Peluso et al., 2009). If this hyperactivation could be reduced by means of probiotics it might lead to a reduction in symptom severity. Thereby it might represent a potential new treatment for affective psychiatric disorders since gut microbiota have been suggested to play a crucial role in these kinds of disorders. Patients with depression disorder have been shown to have increased levels of antibodies in the blood that are secreted as immune response against LPS, suggesting a dysfunction of the epithelial barrier (Maes, Kubera, & Leunis, 2008). Restoring the epithelial barrier function might thus lead to fewer circulating inflammation markers in patients with depression disorder and might have beneficial effects on their symptom severity. Additionally, gut microbiota are involved in the production of metabolites of, for instance, precursors of serotonin. Due to evidence suggesting serotonin deficiency as a possible causal factor in a number of affective disorders, bacteria producing this precursor might be beneficial for this group of disorders (Dinan et al., 2013; Lakhan & Vieira, 2008). Future studies in these patient populations could be set up in order to test the effectiveness of a probiotic treatment for these specific disorders.

References

Alcock, J., Maley, C. C., & Aktipis, C. A. (2014). Is eating behavior manipulated by the gastrointestinal microbiota? Evolutionary pressures and potential mechanisms. *BioEssays, 36*(10), 940–949.

Allin, K. H., Nielsen, T., & Pedersen, O. (2015). Mechanisms in endocrinology: Gut microbiota in patients with type 2 diabetes mellitus. *European Journal of Endocrinology, 172*(4), R167–R177.

Bercik, P., Denou, E., Collins, J., Jackson, W., Lu, J., Jury, J., … & Collins, S. M. (2011). The Intestinal Microbiota Affect Central Levels of Brain-Derived Neurotropic Factor and Behavior in Mice. *Gastroenterology, 141*(2), 599–609.

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., … & Van Der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 34*(6), 623–648.

Bishop, S., Duncan, J., Brett, M., & Lawrence, A. D. (2004). Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. *Nature Neuroscience, 7*(2), 184–188.

Capuron, L., & Miller, A. H. (2011). Immune system to brain signaling: Neuropsychopharmacological implications. *Pharmacology & Therapeutics, 130*(2), 226–238.

Charnay, Y., & Leger, L. (2010). Brain serotonergic circuitries. *Dialogues in Clinical Neuroscience, 12*(4), 471–487.

Chen, X., D'Souza, R., & Hong, S.T. (2013). The role of gut microbiota in the gut-brain axis: current challenges and perspectives. *Protein & Cell,* 4(6), 403–414.

Collins, S. M., Surette, M., & Bercik, P. (2012). The interplay between the intestinal microbiota and the brain. *Nature Reviews Microbiology, 10*(11), 735–742.

Courtney, S. M., Petit, L., Haxby, J. V., & Ungerleider, L. G. (1998). The role of prefrontal cortex in working memory: examining the contents of consciousness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353*(1377), 1819–1828.

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience, 13*(10), 701–712.

Cryan, J. F., & O'Mahony, S. M. (2011). The microbiome-gut-brain axis: from bowel to behavior. *Neurogastroenterology & Motility, 23*(3), 187–192.

Desbonnet, L., Clarke, G., Shanahan, F., Dinan, T. G., & Cryan, J. F. (2014). Microbiota is essential for social development in the mouse. *Molecular Psychiatry, 19*(2), 146–148.

Dinan, T. G., Stanton, C., & Cryan, J. F. (2013). Psychobiotics: A Novel Class of Psychotropic. *Biological Psychiatry, 74*(10), 720–726.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences, 15*(2), 85–93.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving Emotional Conflict: A Role for the Rostral Anterior Cingulate Cortex in Modulating Activity in the Amygdala. *Neuron, 51*(6), 871–882.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cerebral Cortex, 17*(11), 2669–2674.

Foster, J. A., & McVey Neufeld, K. A. (2013). Gut–brain axis: how the microbiome influences anxiety and depression. *Trends in Neurosciences, 36*(5), 305–312.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-Related fMRI: Characterizing Differential Responses. *NeuroImage, 7*(1), 30–40.

Ghosh, S., Van Heel, D., & Playford, R. J. (2004). Probiotics in inflammatory bowel disease: is it all gut flora modulation?. *Gut, 53*(5), 620–622.

Gimeno, D., Kivimäki, M., Brunner, E. J., Elovainio, M., De Vogli, R., Steptoe, A., … & Ferrie, J. E. (2009). Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study. *Psychological*

Medicine, 39(03), 413–423.

Gold, A. L., Morey, R. A., & McCarthy, G. (2015). Amygdala–Prefrontal Cortex Functional Connectivity During Threat-Induced Anxiety and Goal Distraction. *Biological Psychiatry, 77*(4), 394–403.

Hargreaves, K. M. (1990). Neuroendocrine markers of stress. *Anesthesia Progress, 37*(2–3), 99– 105.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport, 11*(1), 43–48.

Haxby, J.V., Hoffman, E.A., & Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.

Joormann, J., & Gotlib, I. H. (2010). Emotion Regulation in Depression: Relation to Cognitive Inhibition. *Cognition and Emotion, 24*(2), 281–298.

Juruena, M. F. (2014). Early-life stress and HPA axis trigger recurrent adulthood depression. *Epilepsy & Behavior, 38*, 148–159.

Kajander, K., Hatakka, K., Poussa, T., Färkkilä, M., & Korpela, R. (2005). A probiotic mixture alleviates symptoms in irritable bowel syndrome patients: a controlled 6-month intervention. *Alimentary Pharmacology & Therapeutics, 22*(5), 387–394.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology. General, 132*(1), 47–70.

King, M. S. (2007). Anatomy of the Rostral Nucleus of the Solitary Tract. In R. M. Bradley (Ed.), *The Role of the Nucleus of the Solitary Tract in Gustatory Processing*. Boca Raton (FL): CRC Press/Taylor & Francis.

Kopp, M. V., Goldstein, M., Dietschek, A., Sofke, J., Heinzmann, A., & Urbanek, R. (2008). Lactobacillus GG has in vitro effects on enhanced interleukin-10 and interferon-gamma release of mononuclear cells but no in vivo effects in supplemented mothers and their neonates. *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology, 38*(4), 602–610.

Lakhan, S. E., & Vieira, K. F. (2008). Nutritional therapies for mental disorders. *Nutrition Journal, 7*, 2.

Logan, A. C., & Katzman, M. (2005). Major depressive disorder: probiotics may be an adjuvant therapy. *Medical Hypotheses, 64*(3), 533–538.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7*(3), 166–174.

Lyte, M. (2014). Microbial endocrinology: Host-microbiota neuroendocrine interactions influencing brain and behavior. *Gut Microbes, 5*(3), 381–389.

Maes, M., Kubera, M., & Leunis, J.C. (2008). The gut-brain barrier in major depression: intestinal mucosal dysfunction with an increased translocation of LPS from gram negative enterobacteria (leaky gut) plays a role in the inflammatory pathophysiology of depression. *Neuroendocrinology Letters, 29*(1), 117–124.

Mayer, E. A. (2011). Gut feelings: the emerging biology of gut–brain communication. *Nature Reviews Neuroscience, 12*(8), 453–466.

Messaoudi, M., Lalonde, R., Violle, N., Javelot, H., Desor, D., Nejdi, A., … & Cazaubiel, J.M. (2011). Assessment of psychotropic-like properties of a probiotic formulation (Lactobacillus helveticus R0052 and Bifidobacterium longum R0175) in rats and human subjects. *The British Journal of Nutrition, 105*(05), 755–764.

Miller, A. H. (1998). Neuroendocrine and immune system interactions in stress and depression. *Psychiatric Clinics of North America, 21*(2), 443–463.

Moayyedi, P., Ford, A. C., Talley, N. J., Cremonini, F., Foxx-Orenstein, A. E., Brandt, L. J., & Quigley, E. M. M. (2010). The efficacy of probiotics in the treatment of irritable bowel syndrome: a systematic review. *Gut, 59*(3), 325–332.

Naseribafrouei, A., Hestad, K., Avershina, E., Sekelja, M., Linløkken, A., Wilson, R., & Rudi, K. (2014). Correlation between the human fecal microbiota and depression. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society, 26*(8), 1155–1162.

Neufeld, K. M., Kang, N., Bienenstock, J., & Foster, J. A. (2011). Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterology & Motility, 23*(3), 255–e119.

Neuman, H., Debelius, J. W., Knight, R., & Koren, O. (2015). Microbial endocrinology: the interplay between the microbiota and the endocrine system. *FEMS Microbiology Reviews, 39*(4), 509–521.

O'Mahony, S. M., Clarke, G., Borre, Y. E., Dinan, T. G., & Cryan, J. F. (2015). Serotonin, tryptophan metabolism and the brain-gut microbiome axis. *Behavioural Brain Research, 277*, 32–48.

O'Mahony, S. M., Marchesi, J. R., Scully, P., Codling, C., Ceolho, A.M., Quigley, E. M. M., … & Dinan, T. G. (2009). Early life stress alters behavior, immunity, and microbiota in rats: implications for irritable bowel syndrome and psychiatric illnesses. *Biological Psychiatry, 65*(3), 263–267.

Peluso, M. A. M., Glahn, D. C., Matsuo, K., Monkul, E. S., Najt, P., Zamarripa, F., … & Soares, J. C. (2009). Amygdala hyperactivation in untreated depressed individuals. *Psychiatry Research, 173*(2), 158–161.

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

Roberts, K. L., & Hall, D. A. (2008). Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory Stroop tasks. *Journal of Cognitive Neuroscience, 20*(6), 1063–1078.

Rook, G. A. W., Lowry, C. A., & Raison, C. L. (2013).

Microbial "Old Friends", immunoregulation and stress resilience. *Evolution, Medicine, and Public Health, 2013*(1), 46–64.

Rothwell, N. J., & Hopkins, S. J. (1995). Cytokines and the nervous system II: Actions and mechanisms of action. *Trends in Neurosciences, 18*(3), 130–136.

Santos, J., Yang, P. C., Söderholm, J. D., Benjamin, M., & Perdue, M. H. (2001). Role of mast cells in chronic stress induced colonic epithelial barrier dysfunction in the rat. *Gut, 48*(5), 630-636.

Sherman, M. P., Zaghouani, H., & Niklas, V. (2015). Gut microbiota, the immune system, and diet influence the neonatal gut-brain axis. *Pediatric Research, 77*(1–2), 127–135.

Shikora, S. A., Toouli, J., Herrera, M. F., Kulseng, B., Brancatisano, R., Kow, L., … & Billingto, C. J. (2015). Intermittent Vagal Nerve Block for Improvements in Obesity, Cardiovascular Risk Factors, and Glycemic Control in Patients with Type 2 Diabetes Mellitus: 2-Year Results of the VBLOC DM2 Study. *Obesity Surgery, 26*(5), 1021–1028.

Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology, 31*(1), 69–75.

Smith, K., McCoy, K. D., & Macpherson, A. J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Seminars in Immunology, 19*(2), 59–69.

Söderholm, J. D., & Perdue, M. H. (2001). II. Stress and intestinal barrier function. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 280*(1), G7-G13.

Sparkman, N. L., Buchanan, J. B., Heyen, J. R. R., Chen, J., Beverly, J. L., & Johnson, R. W. (2006). Interleukin-6 Facilitates Lipopolysaccharide-Induced Disruption in Working Memory and Expression of Other Proinflammatory Cytokines in Hippocampal Neuronal Cell Layers. *The Journal of Neuroscience, 26*(42), 10709–10716.

Steenbergen, L., Sellaro, R., Van Hemert, S., Bosch, J. A., & Colzato, L. S. (2015). A randomized controlled trial to test the effect of multispecies probiotics on cognitive reactivity to sad mood. *Brain, Behavior, and Immunity, 48,* 258–264.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662.

Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., Yu, X.N., … & Koga, Y. (2004). Postnatal microbial colonization programs the hypothalamic–pituitary–adrenal system for stress response in mice. *The Journal of Physiology, 558*(1), 263–275.

Swick, D., Ashley, V., & Turken, A. U. (2008). Left inferior frontal gyrus is critical for response inhibition. *BMC Neuroscience, 9*(1), 102.

Tillisch, K., Labus, J., Kilpatrick, L., Jiang, Z., Stains, J., Ebrat, B., … & Mayer, E. A. (2013). Consumption of Fermented Milk Product With Probiotic Modulates Brain Activity. *Gastroenterology, 144*(7), 1394–1401.

Van Hemert, S., & Ormel, G. (2014). Influence of the Multispecies Probiotic Ecologic®Barrier on Parameters of Intestinal Barrier Function. *Food and Nutrition Sciences, 5*(18), 1739–1745.

Winecoff, A., Clithero, J. A., Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. A. (2013). Ventromedial Prefrontal Cortex Encodes Emotional Value. *Journal of Neuroscience, 33*(27), 11032–11039.

Wu, H.J., & Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes, 3*(1), 4–14.

Zoccatelli, G., Beltramello, A., Alessandrini, F., Pizzini, F. B., & Tassinari, G. (2010). Word and position interference in Stroop tasks: a behavioral and fMRI study. *Experimental Brain Research, 207*(1–2), 139–147.

The Effects of Grammatical Gender on Reference Processing in German: an ERP Study

Julia Misersky[4]
Supervisors: Asifa Majid[1,3,4], Tineke M. Snijders[2,4], Pascal M. Gygax[5]

[1]*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*
[2]*Radboud University Nijmegen, Donders Centre for Cognitive Neuroimaging, The Netherlands*
[3]*Radboud University Nijmegen, Centre for Language Studies, The Netherlands*
[4]*Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands*
[5]*Université de Fribourg, Fribourg, Switzerland*

Grammatically masculine forms can refer to men and women, but may favour male-specific interpretations. Using a match-mismatch paradigm, the present event-related potential (ERP) study assessed how masculine grammatical gender in role nouns affects referent processing in German. Twenty participants read sentences in which a stereotypically neutral role noun (grammatically masculine or feminine) introduced a group of people. A sentence continuation specified the group as consisting in part of *men* or *women*, meaning continuations were either congruent (masculine–*men*, feminine–*women*) or incongruent (masculine–*women*, feminine–*men*) to the grammatical gender of the role noun. Incongruent continuations were expected to result in an N400-P600 complex. Between 300 and 500 ms, no N400-like effect was observed. Following masculine role nouns, all continuations were processed similarly (p = .891). Following feminine role nouns, incongruent continuations elicited more positive responses than congruent ones (p = .045). Between 500 and 800 ms, a P600-like effect was observed. For both masculine and feminine role nouns, incongruent continuations resulted in more positive responses than congruent ones (p = .039). The results are discussed in terms of a two-stage model: initially, the incongruency between the masculine and *women* continuations goes unnoticed, yet leads to processing difficulties later on, implying a male-specific interpretation.

Keywords: grammatical gender, reference processing, EEG, P600, N400

Corresponding author: Julia Misersky; **E-mail:** julia.misersky@mpi.nl

Gender is a social category we encounter on a daily basis, with gender equality having become an important socio-political issue. Cues to gender are present in language in many ways. For example, we find words describing gendered referents (e.g., *father* or *woman*). Language can also carry stereotype information, which is strongly linked to conceptual knowledge (e.g., *electrician* or *nurse*, associated with males and females, respectively). Additionally, grammatical gender may map onto social categories of male and female gender. Often, we find these mappings in so-called role nouns, which describe social importance or occupation, and constitute an important way of denoting human referents. Indeed, grammatical gender in role nouns has been considered increasingly relevant to the equality debate. It has been suggested that masculine grammatical gender in role nouns favours male-specific interpretations, yet the masculine is regularly used to generically talk about male and female referents. This raises the question of whether the masculine is adequate in representing society in a gender-neutral way. To assess this issue, the present study investigated how masculine grammatical gender affects our understanding of differently gendered referents.

Behavioural research on gender cues

As a linguistic category, gender has been described to be "almost universally present in language […]" (Irmen, Holt, & Weisbrod, 2010, p. 133). Indeed, the gender of a human referent can be indicated linguistically via diverse means. While lexical semantics (e.g., *woman*) cue gender explicitly, role nouns (e.g., *musician*) often only imply referent gender. Take the example of the role noun *musician(s)*. In English, the noun is not grammatically gendered in the singular or plural. Only via the use of pronouns, and only in the singular can the gender of the *musician* be revealed as she or he. In contrast, in many grammatically gendered languages, the same role noun is morpho-syntactically marked for gender. It seems, then, that role nouns are realised differently across languages, depending on the grammatical system.

Behavioural research across languages has investigated how role nouns are understood. In the absence of grammatical gender, referent gender is cued by stereotypicality (Gygax, Gabriel, Sarrasin, Oakhill, & Garnham, 2008; Sato, Gygax, & Gabriel, 2013). For example, in English, the noun *musicians* implies male referents (Misersky et al., 2014). In gendered languages, such as German, grammatical gender indicates referent gender, since specific forms are used to describe females and males. While feminine role nouns (e.g., $Musikerinnen_{\text{feminine}}$ [fem.]) cue specifically for females, the masculine (e.g., $Musiker_{\text{masculine}}$ [masc.]) can be used for males, but it can be used as a generic too. As such, $Musiker_{\text{masc.}}$ can be used for a group of men or a mixed-gender group consisting of both men and women. In short: the masculine may yield specific as well as generic interpretations.

Behavioural research has revealed using a specific form such as $Musiker_{\text{masc.}}$ for a mixed-gender group (i.e., including both males and females) results in an ambiguity in interpretation. For example, Gygax et al. (2008) assessed the gender interpretation of role nouns by comparing English, which has natural gender, to French and German, both of which have grammatically masculine and feminine gender categories. Participants were presented with a stereotypically male or female role noun (in the masculine for French and German) in one sentence, and with an anaphoric noun (*men* or *women*) in a second. With a Yes or No response, participants had to indicate whether the sentence containing the anaphor was a sensible continuation of the sentence with the role noun. In English, participants' responses and response times were linked to the stereotypicality of the role noun. There was no such effect of stereotypicality for French and German. Instead, in both these languages *men* continuations received more Yes responses (i.e., continuation is sensible) compared to *women* continuations. This was further reflected in the response times: participants were faster to judge *men* continuations as sensible compared to *women* continuations. Thus, despite being generically intended, the masculine was interpreted as specific. Recent studies have found similar effects in primary school children (Vervecken, Gygax, Gabriel, Guillod, & Hannover, 2015; Vervecken & Hannover, 2015). Presenting role nouns in the masculine or a gender-neutral form in German and Dutch, grammatical form was found to affect ratings of job status, difficulty, and accessibility. Specifically, the children reported lower self-efficacy –that is, whether they thought they could succeed in a job– when presented with the masculine. These results highlight the societal relevance of grammatical gender cues.

In sum, this research suggests grammatical gender to be highly relevant in guiding our understanding of human referents, which may even override stereotype information (Gygax et al., 2008; Irmen & Roßberg, 2004). In particular, grammatically masculine forms

seem to lead to male-specific interpretations, despite their use as a generic for all genders.

ERP research on gender cues

Electroencephalography (EEG), in particular event-related-potentials (ERPs), has been used to assess the effects of gender cues on comprehension during online processing. ERP studies on gender cues regularly find N400 and P600 effects. A negative deflection around 300 to 500 ms after stimulus presentation, the N400 is generally linked to lexical-semantic processing (Kutas & Federmeier, 2011). The P600 is sensitive to syntactic mismatch and integration difficulties (Osterhout, McLaughlin, & Bersick, 1997b), and is characterised by a positive deflection around 500 to 800 ms after stimulus presentation.

To a large extent, ERP research on the effects of grammatical gender has studied the processing of non-human referents, such as object nouns. Recent work investigated the processing of gender-to-ending consistency in Italian (Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani, & Cacciari, 2015). The materials consisted of sentences containing determiner-noun pairs, which either matched in gender (*il*$_{masc.}$ *cucchiaio*$_{masc.}$ [the spoon]) or did not (*la*$_{fem.}$ *cucchiaio*$_{masc.}$ [the spoon]). Gender mismatches resulted in processing difficulties reflected by a P600 effect. Similar results were obtained with noun-adjective gender mismatches (e.g., *faro*$_{masc.}$*–alta*$_{fem.}$ [lighthouse–high]) in Spanish sentences (Barber & Carreiras, 2005). Interestingly, the authors found the same mismatches presented as word-pairs outside of sentences elicited an N400 effect. According to Barber and Carreiras (2005), the N400 component can be driven by conceptual, lexical or morphological feature integration depending on task and stimuli. The authors thus concluded the N400 for the word-pairs reflected simple feature integration. By contrast, full sentences required more complex syntactic structure building, leading to the P600.

ERP research on human referent processing has predominantly focused on stereotype information embedded in role nouns. White, Crites, Taylor, and Corral (2009) used a match-mismatch paradigm and found word-pairs, which mismatched in stereotypicality (e.g., *secretary–aggressive*) elicited a larger N400 effect than matching word-pairs (e.g., *secretary–caring*). Research by Osterhout, Bersick, and McLaughlin (1997a) assessed the processing of reflexive pronouns (*himself/herself*) following a role noun in English. Pronouns either did or did not match an antecedent role noun's gender definition (e.g., *bachelor*) or stereotype (e.g., *doctor*). The results revealed a P600 effect for gender mismatches. Interestingly, the amplitude of the P600 varied for stereotypical compared to definitional nouns. Mismatches between reflexives and definitional role nouns were considered more anomalous as they leave little room for interpretation, thus leading to larger P600 amplitude. Interestingly, an N400 was also found for the last word of each sentence for the definition-violating reflexives. The authors explain that in the stereotype-violating sentences, an acceptable interpretation had been derived. For the definition-violating sentences, this was not possible, leading to the N400. Testing German speakers, Irmen et al. (2010) investigated how stereotypicality of role nouns affects the processing of a subsequent referent. They presented their participants with sentences consisting of a stereotypically male or female role noun (e.g., *computer scientist*; stereotypically male), and a co-referential continuation, which was either neutral (e.g., *these people*), matching (e.g., *these men*) or mismatching (e.g., *these women*) with regards to gender. An N400 effect across all continuations following a stereotypically male role was observed, suggesting participants may have anticipated a mismatch. A later P600 effect shows a clear interaction between the stereotypicality of the role noun and the continuation, and was taken as the integration of the two nouns. Irmen et al. (2010) linked this to a two-stage model of reference resolution (Garrod & Terras, 2000): initial linking (bonding) relies on lexical-semantic information, whereas resolution takes place once additional information has been taken into account.

In sum, two ERP components are most regularly observed in the research on gender cues: the N400 and P600. Roughly speaking, the N400 reflects lexical-semantic processes, whereas the P600 is associated with systematic syntactic processing. Additionally, an N400-P600 complex has been described (e.g., Irmen et al., 2010), which might be especially relevant to the processing of role nouns. Both conceptual and syntactic information are relevant to successfully building co-reference (Schmitt, Lamers, & Münte, 2002), and this might be especially true for languages where semantics are increasingly subject to syntactic constraints as a result of grammatical gender. In German, for example, the semantic and grammatical gender of words describing human referents (e.g., *Frauen*$_{fem.}$ [women]) tend to agree (Irmen et al., 2010). This means lexical-semantic and syntactic processing difficulties may co-occur during the processing of grammatical gender cues for human referents. In line with previous interpretations of

the two ERP components, initial reference might be established by the bonding of conceptual and/or morphological features (N400), with a systematic analysis of grammatical gender taking place later on (P600).

The present study

Behavioural research shows how the grammatical gender of role nouns can affect understanding of human referents. Following a feminine form, only female referents are pragmatically and grammatically acceptable. By contrast, it is pragmatically acceptable to use a masculine form for male and female referents. Grammatically, however, only male referents are an acceptable match for a masculine role noun. The masculine has thus been suggested to lead to male-specific interpretations (Gygax et al., 2008), at least when explicit decision-making is required. ERPs provide the ideal measure to assess referent processing as separate from decision-making. Thus far, ERP research on grammatical gender has predominantly studied processing of nonhuman referents (e.g., Caffarra et al., 2015). In the case of human referent processing, ERP studies using role nouns have focused on stereotype processing (e.g., White et al., 2009). Motivated by Gygax et al.'s (2008) findings, the present study investigated how masculine grammatical gender affects the processing of differently gendered human referents. Specifically, the study aimed to assess whether the masculine is understood as gender-neutral (i.e., encompassing both females and males) or as specific to males. Focusing on the N400 and P600, which relate to semantic and syntactic processing respectively, allowed for a comprehensive investigation of this question.

Grammatical gender may override stereotype information in decision-making, but the two are considered to interact during processing (Gygax et al., 2008; Irmen et al., 2010). In a match-mismatch paradigm, the present study thus used stereotypically neutral roles to systematically focus solely on the effects of grammatical gender. The experimental sentences introduced a group of people via a role noun (manipulated as grammatically masculine or feminine, e.g., *Studenten*$_{masc.}$ or *Studentinnen*$_{fem.}$), and a sentence continuation specified the group as consisting in part, yet not exclusively, of men (*Männer*$_{masc.}$) or women (*Frauen*$_{fem.}$). This meant role noun and continuation either matched (masculine–*men*; feminine–*women*) or mismatched (masculine–*women*; feminine–*men*) in grammatical gender. Table 1 gives an example of the sentence stimuli.

Both the N400 and P600 have been observed in ERP studies on role nouns, and both conceptual and syntactic processes have been considered important for reference building. Thus, in this study, both components resulting from grammatical gender incongruencies between role nouns and continuations were expected. The analyses specifically focused on ERPs following the onset of the continuations (*men, women*) to assess how these are integrated with the preceding role noun. In line with the main research question, the following findings were hypothesised:

Firstly, it was anticipated that a role noun in the masculine would favour a male-specific interpretation of the referent. As a result, masculine role nouns

Table 1.

Example of the sentence stimuli and conditions.

Grammatical gender	Continuation	
Masculine	congruent (*men*)	Die **Studenten**$_{masculine}$ gingen zur Mensa, weil manche der **Männer** Hunger hatten. [The **students** went to the canteen, because some of the **men** were hungry.]
	incongruent (*women*)	Die **Studenten**$_{masculine}$ gingen zur Mensa, weil manche der **Frauen** Hunger hatten. [The **students** went to the canteen, because some of the **women** were hungry.]
Feminine	congruent (*women*)	Die **Studentinnen**$_{feminine}$ gingen zur Mensa, weil manche der **Frauen** Hunger hatten. [The **students** went to the canteen, because some of the **women** were hungry.]
	incongruent (*men*)	Die **Studentinnen**$_{feminine}$ gingen zur Mensa, weil manche der **Männer** Hunger hatten. [The **students** went to the canteen, because some of the **men** were hungry.]

followed by *women* continuations were hypothesised to lead to processing difficulties. In particular, both an N400 and a P600 effect were expected to reflect these difficulties, with incongruencies leading to a relatively more negative amplitude of the N400, and a relatively more positive amplitude of the P600.

Secondly, the masculine is used as a default for both male and female referents, whereas the feminine is only used for female referents. Theoretically, this allows for a more flexible gender interpretation following masculine role nouns. Processing difficulties (as reflected by the N400 and P600) were thus expected to be reduced for the masculine (masculine–*women*), compared to the feminine role nouns (feminine–*men*).

Gygax and Gabriel (2008) suggest using materials with both feminine and masculine forms may lead participants to consider them in direct contrast. This may amplify the interpretation of the masculine as specific to males, since the feminine is always specific to females. However, the motivation for including the feminine in this study was twofold. Firstly, since role nouns in the feminine always denote female referents, a pairing with male referents would be highly incongruent. The high specificity of the feminine would establish a benchmark of incongruency between role noun and continuation. This would allow for a more comprehensive interpretation of the results concerning continuations after a masculine form, which has a flexible use for all genders. Secondly, we often encounter both forms in a variety of contexts. As such, including both forms in this study provided an appropriate representation of a real-life situation.

Methods

Participants

Twenty native speakers of German (13 female, age range 19 – 29 years, M = 22.3, SD = 2.68) recruited from Radboud University's SONA system participated in this study. Participants provided written informed consent, and received course credit or payment for their participation. All had normal or corrected to normal vision and were right-handed. Four participants were excluded from further analysis; two due to a low number of trials after pre-processing, and another two due to accuracy on the content questions (described below) that was at or below chance. The study was approved by the local ethical committee (Commissie Mensgebonden Onderzoek, Regio Arnhem-Nijmegen).

Materials and design

A total of 156 role nouns were selected from a recent norming study (Misersky et al., 2014) on the basis of their stereotypicality rating. Role nouns rated as stereotypically neutral (M = .47, SD = .08, ranging between 1 = stereotypically female and 0 = stereotypically male) were included in the materials. Using these role nouns at the beginning of the sentence, coordinate clauses were created inspired by previous research (Gygax et al., 2008; Irmen et al., 2010). The sentence-initial role noun introduced a group of people, with the role nouns grammatical gender manipulated as masculine (e.g., *Studenten*$_{masc.}$) or feminine (e.g., *Studentinnen*$_{fem.}$). Later in the sentence, this group was specified as being partially consisting of men (*Männer*$_{masc.}$) or women (*Frauen*$_{fem.}$). Inspired by previous research (Gygax et al., 2008), pseudo-randomised quantifiers were selected to highlight the group was not exclusively made up of men or women; *einige* (a few/some), *mehrere* (several), *manche* (some), *einzelne* (single ones) or *viele* (many); refer to Table 1 for an example. In addition, 80 filler items were created, half of which followed a similar format, and half of which were structurally different, resulting in a total of 236 experimental sentences. Sentences were pseudo-randomised for presentation.

Procedure

Participants were seated in a dimly illuminated sound-attenuating testing booth. They were instructed to attentively read the sentences, since they would receive content questions throughout the experiment. These instructions were both explained by the experimenter, and presented visually on the testing PC. Since eye movements distort the EEG recording, participants were also asked to only blink between sentences and during breaks. Participants were able to remain in contact with the experimenter via microphone. All interactions and instructions were in German.

Sentences were presented using Presentation software (Neurobehavioral Systems, www.neurobs.com). Each sentence was presented in a word-by-word serial visual presentation mode at the centre of a 24-inch PC monitor. The background was a dark grey with the words presented in white letters (Helvetica, font size 26). The beginning of each sentence was preceded by a fixation cross (+). Each word was flashed for 380 ms with an inter-

word-interval (IWI) of 145 ms. The second and fourth word of each sentence were flashed slightly longer, for 480 ms, due to word length. Sentence-final words were followed by a full stop, then a 1000 ms blank. Every ten sentences, a content question would appear on-screen, requiring a self-paced Yes or No response via button press with the left or right index finger, respectively. The question related to the activity carried out in the sentence; and there was no repetition of the role noun. The inter-trial-interval (ITI) was 2000 ms during which the fixation cross re-appeared. First, participants received nine practice sentences, after which remaining uncertainties about the task could be resolved. The experiment was split into four blocks of 59 trials each. There were self-paced pauses between blocks where a drink of water was offered to the participants. Figure 1 gives an example of the procedure.

EEG set-up and apparatus

Continuous EEG was recorded from 32 active electrodes (10-20 system) attached to an elastic cap (actiCAP), with a BrainAmp DC amplifier (Brain Products, Gilching, Germany). The signal was sampled at 500 Hz. One electrode in the cap provided an active ground. Electrooculogram (EOG) was recorded from electrodes above and below the eye, and at the outer canthi of the eyes. Electrode impedances were kept below 20 kΩ.

The data was pre-processed using the FieldTrip toolbox for EEG/MEG-analysis (www.fieldtriptoolbox.org; Oostenveld, Fries, Maris, & Schoffelen, 2011) in MATLAB. For each continuation noun, segments were chosen in the range from 200 ms before to 1000 ms after word onset. Offline-filtering included a low-pass filter at 35 Hz and a high-pass filter at 0.1 Hz. The

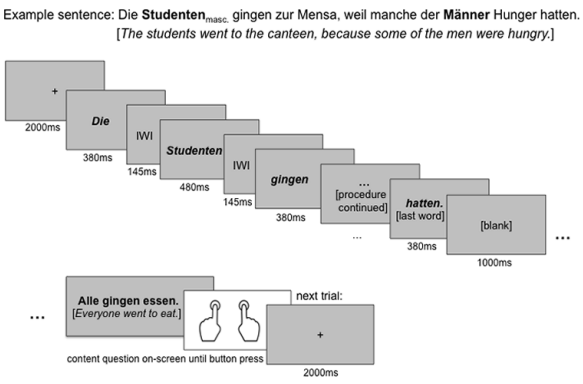

Fig. 1. Example of the word-by-word stimulus presentation followed by a content question.

data were then inspected visually, and trials showing electrode jumps and/or drifting were removed in preparation for an independent component analysis (ICA). ICA was performed to remove any remaining EOG and/or electrocardiogram (ECG) artefacts from the data. All channels were then referenced via the average of the signal of both mastoids (Luck, 2005). A baseline correction was applied in which the signal was normalised relative to a 200 ms stimulus preceding window. Trials containing signal exceeding ±75 μV were removed, and mean ERP amplitudes for the time windows of interest were calculated. The datasets of two subjects were excluded from further analysis, since less than 29 trials per condition (< 25 percent) remained after pre-processing. The average amount of kept trials per condition for the included subjects was 34.65 (88.85%, ranging from an average of 34.43 to 34.75 trials across all conditions).

Analysis

In line with existing work (Irmen et al., 2010; Osterhout et al., 1997a), mean ERP amplitudes were statistically analysed in two main time windows after the onset of the continuation noun; 300 to 500 ms for the N400, and 500 to 800 ms for the P600, respectively. The mean ERP amplitudes were analysed in SPSS. As in Irmen et al. (2010), nine electrodes in anterior, central and posterior positions of the left and right hemisphere and the midline were used for the statistical analyses (F3/z/4, C3/z/4, P3/z/4).

Results

Responses to content questions

The participants were accurate in correctly answering the content questions ($M_{\text{present correct}}$ = 99.18, *SD* = 3.26), meaning they understood the task and were attentively reading the sentences throughout the experiment.

Event-related potentials

Following Irmen et al. (2010), the mean amplitudes of the ERPs for the time windows of interest were subjected to repeated-measures analyses of variance (ANOVA). The factors submitted to each ANOVA included Anteriority (three levels: anterior, central, posterior), Laterality (three levels: left, midline, right), Grammatical Gender of the role noun (two levels:

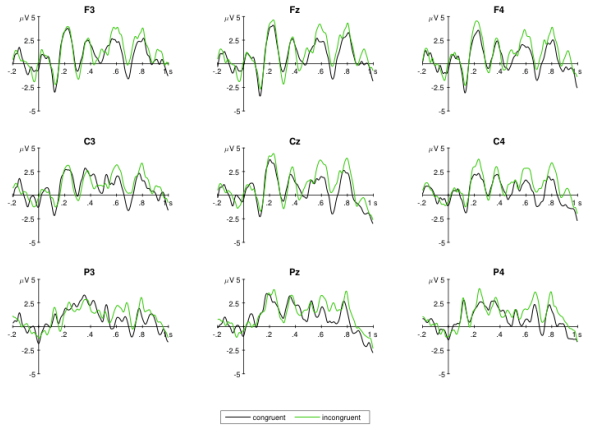

Fig. 2. ERPs for Continuations following a role noun with masculine grammatical gender.

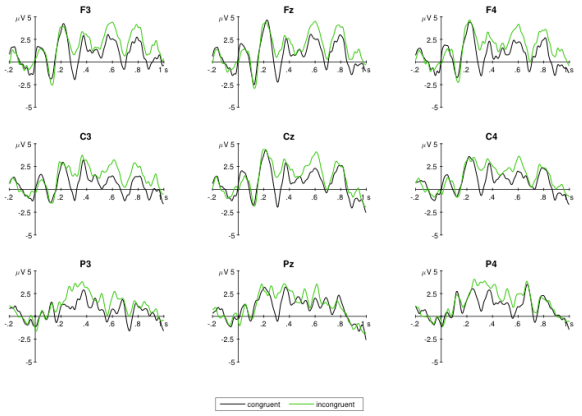

Fig. 3. ERPs for Continuations following a role noun with feminine grammatical gender.

masculine, feminine, see Table 1), and Continuation (two levels: congruent, incongruent, see Table 1). An alpha level of .05 was used for all statistical tests. Note that the effects for Grammatical Gender, Continuation, and the interaction between them are of prime relevance with regards to assessing the effects of grammatical gender on referent processing. Figures 2 and 3 represent the grand average ERPs for the nine electrodes separated by Grammatical Gender.

300 to 500 ms time window. The ANOVA revealed an interaction between Grammatical Gender and Continuation, $F(1, 15) = 5.82$, $p = .029$, $\eta^2 = .28$. Follow-up analyses were carried out for each Grammatical Gender separately (Fig. 4). For role nouns with masculine Grammatical Gender, there was no significant difference between congruent ($M = 1.23$ µV, $SEM = .42$) and incongruent ($M = 1.18$ µV, $SEM = .36$) continuations, $F(1, 15) = .02$, $p = .891$, $\eta^2 = .001$. For role nouns with feminine grammatical gender, however, responses to congruent Continuations differed significantly from those to incongruent ones, $F(1, 15) = 4.78$, $p = .045$, $\eta^2 = .24$. Incongruent Continuations elicited more positive responses ($M = 2.29$ µV, $SEM = .44$) compared to congruent Continuations ($M = 1.18$ µV, $SEM = .47$). There were no main effects of Grammatical Gender or Continuation, nor of Anteriority or Laterality.

500 to 800 ms time window. The ANOVA showed a main effect of Continuation, $F(1, 15) = 5.13$, $p = .039$, $\eta^2 = .25$. Regardless of the Grammatical Gender of the role noun, incongruent Continuations elicited significantly more positive responses ($M = 2.09$ µV, $SEM = .46$), compared to congruent Continuations ($M = 1.07$ µV, $SEM = .32$, see Fig. 5). In this time window, there were no main effects of Anteriority or Laterality, nor any interactions between any of the factors.

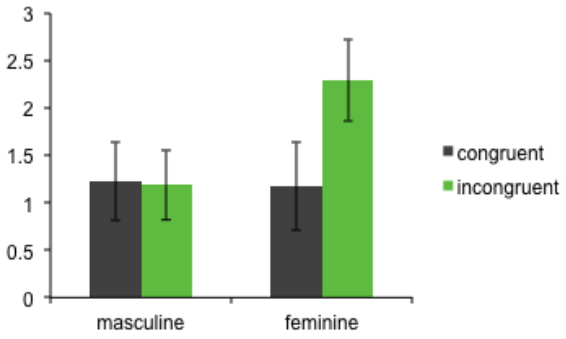

Fig. 4. Mean amplitudes in the 300–500 ms time window as a factor of Grammatical Gender (masculine vs. feminine) and Continuation (men vs. women).

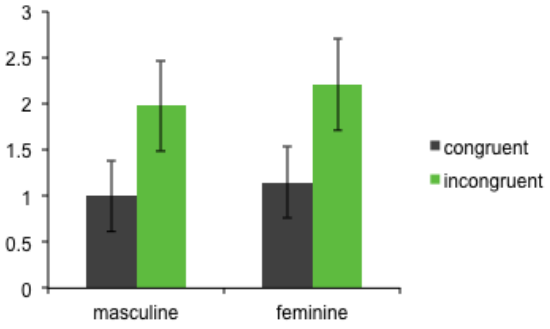

Fig. 5. Mean amplitudes in the 500–800 ms time window as a factor of Grammatical Gender (masculine vs. feminine) and Continuation (men vs. women).

Discussion

The present study assessed how grammatical gender affects referent processing. More specifically, the study focused on how grammatically masculine forms affect the processing of differently gendered referents. In line with previous research, masculine role nouns followed by *women* continuations were expected to lead to processing difficulties reflected by a complex of N400 (300 to 500 ms) and P600 (500 to 800 ms). Furthermore, effects of incongruency were expected to be reduced for continuations following a grammatically masculine role noun (masculine–*women*) compared to a grammatically feminine role noun (feminine–*men*).

In the 300 to 500 ms time window, an interaction between grammatical gender and continuation was found. Put differently, how the continuation was processed depended on the preceding grammatical gender. When participants read a role noun in the masculine, they processed both congruent and incongruent continuations similarly. By contrast, when they read a role noun in the feminine, the continuations were processed differently. Interestingly, the incongruent continuations elicited more positive responses compared to the congruent ones. This was unexpected, since for the N400, incongruencies tend to elicit relatively more negative responses compared to congruencies. Siyanova-Chanturia, Pesciarelli and Cacciari (2012) have looked into the processing of pronouns (*lei*[she]/*lui*[he]) following stereotypically (e.g., *insegnante*$_{\text{masc./fem}}$ [teacher]), and grammatically gendered (e.g., *pensionato*$_{\text{masc.}}$ [pensioner]) role nouns in Italian. For female participants, they observed positive responses to incongruencies between grammatically gendered roles and pronouns, and interpreted this effect as P300-like. The P300 has been linked to stimuli evaluation (Kutas & Hillyard, 1980a; Kutas & Hillyard, 1980b), and is sensitive to this evaluation being task-relevant (Holcomb, 1988). In line with this, Siyanova-Chanturia et al. (2012) attribute their finding to participants having made decisions on the word-pairs during the experiment. Unlike Siyanova-Chanturia et al. (2012), the present study used sentence stimuli. Additionally, participants did not need to explicitly evaluate the continuations, yet answering content questions may have increased the likelihood of participants evaluating the words more closely. While this opens up the possibility for a P300-like effect, additional exploration is needed to sufficiently examine this unexpected effect.

In the 500 to 800 ms time window, a relatively more positive response to incongruent compared to congruent continuations was observed, regardless of the grammatical gender of the role noun preceding it. This result reflects a P600-like effect, meaning participants encountered processing difficulties upon reading *women* when preceded by a grammatically masculine role noun, and upon reading *men* when preceded by a grammatically feminine role noun.

In line with Irmen et al. (2010), the results can be linked to Garrod & Terras' (2000) two-stage reference processing model. During an initial bonding stage, the Continuation is linked to the role noun by an automatic process. Note that complete congruency required the grammatical gender of the role noun to match both the grammatical and lexical-semantic gender of the continuation. Despite the unexpected positive ERP amplitude of the incongruent continuations following a feminine role noun, it was clear bonding was different for the congruent compared to the incongruent continuations. This could be due to the feminine form being highly specific, effectively constraining processing early on. For role nouns in the masculine, however, all continuations were processed similarly. The absence of differential processing in this initial stage could be a result of the masculine form being pragmatically used to describe mixed-gender groups. Only in the later time window, during the reference resolution stage, incongruent continuations lead to processing difficulties with no difference between the two grammatical genders.

To sum up, the observed pattern of effects suggests the grammatical gender of the role noun guided the processing of the continuation. Firstly, a role noun in the feminine leads to differential processing for *men* and *women* continuations early on. Secondly, the mismatch between *women* continuations following a masculine goes unnoticed during initial processing. Thirdly, and importantly, this same mismatch leads to difficulties later on. Linking back to the research question, the results suggest people struggled integrating female referents (*women*) with the grammatically masculine form during late systematic processing. Thus, the grammatically masculine form seems to create a bias towards males, despite being used to encompass females and males. This finding implies the masculine form to be inadequate in representing the whole of society. A grammatical form that is truly gender-neutral and/or encompasses males and females specifically could be key to reducing the effects observed in behavioural research and in this study. A follow-up including such grammatically gender-neutral forms can assess the possibilities of reducing the bias encountered.

Conclusion

The present study has shown that difficulties during referent processing are subject to grammatical constraints. In line with the behavioural research in this field, the data indicate grammatical gender to be an important cue to the understanding of human referents. This has relevant implications for the regular use of the masculine to talk about mixed-gender groups. Since its interpretation favours male referents, future research should aim to assess which grammatical forms can most effectively reduce the processing difficulties observed in this study.

References

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137-153.

Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., & Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology, 52*(8), 1019-1030.

Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*(4), 526-544.

Gygax, P., & Gabriel, U. (2008). Can a group of musicians be composed of women? Generic interpretation of French masculine role names in the absence and presence of feminine forms. *Swiss Journal of Psychology, 67*(3), 143-151.

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes, 23*(3), 464-485.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*(1), 66-85.

Irmen, L., Holt, D. V., & Weisbrod, M. (2010). Effects of role typicality on processing person information in German: Evidence from an ERP study. *Brain Research, 1353*, 133-144.

Irmen, L., & Roßberg, N. (2004). Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology, 23*(3), 272-307.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of Psychology, 62*, 621.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99-116.

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, US: The MIT press.

Misersky, J., Gygax, P. M., Canal, P., Gabriel, U., Garnham, A., Braun, F., ... & Sczesny, S. (2014). Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods, 46*(3), 841-871.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 9 pages.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997a). Brain potentials reflect incongruencies of gender stereotypes. *Memory & Cognition, 25*(3), 273-285.

Osterhout, L., McLaughlin, J., & Bersick, M. (1997b). Event-related brain potentials and human language. *Trends in Cognitive Sciences, 1*(6), 203-209.

Sato, S., Gygax, P. M., & Gabriel, U. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism: Language and Cognition, 16*(4), 792-807.

Schmitt, B. M., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research, 14*(3), 333-346.

Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PloS One, 7*(12), e48712.

Vervecken, D., Gygax, P. M., Gabriel, U., Guillod, M., & Hannover, B. (2015). Warm-hearted businessmen, competitive housewives? Effects of gender-fair language on adolescents' perceptions of occupations. *Frontiers in Psychology, 6*, 1-10.

Vervecken, D., & Hannover, B. (2015). Yes I can! Effects of gender fair job descriptions on children's perceptions of job status, job difficulty, and vocational self-efficacy. *Social Psychology, 46*(2), 76-92.

White, K. R., Crites, S. L., Taylor, J. H., & Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience, 4*, 191-198.