

Maastricht Student Journal  
of  
Psychology and Neuroscience

## Maastricht Student Journal of Psychology and Neuroscience (MSJPN)

### Editors

Anke Sambeth, PhD  
Department of Neuropsychology and Psychopharmacology  
Faculty of Psychology and Neuroscience  
Maastricht University  
Universiteitssingel 40, 6229 ER, Maastricht, The Netherlands  
Tel: 0031 43 388 1757  
anke.sambeth@maastrichtuniversity.nl

Silke Conen, PhD  
Neuroscience and Psychiatry Unit  
University of Manchester  
Stopford Building, Oxford Road, M13 9PT, Manchester, United Kingdom  
Tel: 0044 161 275 1725  
silke.conen@manchester.ac.uk

Tim Leufkens, PhD  
Brain, Body and Behavior Group  
Information and Cognition  
Philips Research  
High Tech Campus 34, 5656 AE, Eindhoven, The Netherlands  
Tel: 0031 6 316 282 20  
tim.leufkens@philips.com

Luís Miguel Tojo, BSc  
Faculty of Psychology and Neuroscience  
Maastricht University  
Universiteitssingel 40, 6229 ER, Maastricht, The Netherlands  
Tel: 0031 6 128 383 61  
l.tojo@student.maastrichtuniversity.nl

Contact email for all MSJPN related material and submission of manuscripts:  
studentjournal-fpn@maastrichtuniversity.nl

### Reviewing board

Sylvia Kreutzer, BSc	Nadia Fischer, BSc
Tahnee Engelen, BSc	Stephan Bruggeman, BSc
Lisa Wittenhagen, BSc	Leon Skottnik, BSc
Dimitra Kostoglou, BSc	Nicolai Groeger, BSc
Marianne Drabek, BSc	Sacha Leinders, BSc

### Disclaimer

The authors, editors, and publisher will not accept any legal responsibility for any errors or omissions that may be made in this publication. The articles reflect only the authors' view.

Lay-out & cover: Joris van Veghel, jorisvanveghe@gmail.com  
Prints: Wöhrmann Print Service, Zutphen, The Netherlands

# Maastricht Student Journal of Psychology and Neuroscience

*Volume 2 Number 1 July 2013*

*Anke Sambeth  
Silke Conen  
Tim Leufkens  
Luís Miguel Tojo  
(eds.)*



## CONTENTS

The Reliability of a Child as an Eyewitness in Court <i>Ingeborg Close</i>	9
The effect of shared leadership on team performance in international undergraduate students <i>Kimani Michalke</i>	15
The Effects of Mindfulness Training on Children's Attention Skills <i>Alice Wellum</i>	28
Driving home for Christmas: Influences of music tempo and inhibition training on simulated driving performance <i>Chris Makkinje</i>	46
Time your stress if you aim for success: the influence of stress on memory for different memory phases <i>Floor van den Brand</i>	58
Fear to get near: personal space in individuals with psychopathic traits <i>Martijn van Teffelen</i>	70
The effects of mindfulness versus thought suppression instruction on the appraisal of emotional and neutral pictures <i>Marko Radivojevic</i>	84
Functional and anatomical connectivity in Complex Regional Pain Syndrome: A Multi-modal magnetic resonance imaging study <i>Emma Biggs</i>	98
Cabbage, Cars and Beer? An examination of Dual Attitudes towards the Concept "German" <i>Sabine Lenz</i>	116
To what extent do the various measures of confidence affect the accuracy-confidence relationship in earwitnesses? A review of research on the earwitness testimony <i>Annette Verhaeg</i>	130
What's the problem with free will? <i>Marianne Drabek</i>	146
'Walking through the hospital', priming tolerance to enhance the approaching behavior between healthy individuals and patients suffering from psychoses <i>Lisa Hubrach &amp; Kiki Nap</i>	155



## LETTER FROM THE EDITORIAL BOARD

Dear reader,

The editorial board of the Maastricht Student Journal of Psychology and Neuroscience (MSJPN) gladly introduces you its second edition. Last year, when we launched the first edition of MSJPN, we certainly had plans to continue providing an annual scientific platform for students at the Faculty of Psychology and Neuroscience (FPN) of Maastricht University. We were delighted to receive so many enthusiastic responses from students, employers, and external relations. We feel that already in one year MSJPN has gained a lot of support by others and that its name and its purpose are becoming increasingly known among all members of FPN.

Yet, we simply had to wait and see whether this year there would again be a group of students that was willing to spend a considerable amount of spare time in writing and reviewing scientific papers. Fortunately, there was. It is with great pleasure that we announce that the current edition of MSJPN comprises a total of 12 papers. Of those 12 papers, eight are original research articles and four are literature reviews. It is absolutely ravishing to conclude that students are submitting papers based, for instance, on research they conducted during their second year's research course. In addition, it has become clear that a Bachelor's thesis creates a perfect starting point to eventually publish an impressive review article in this journal.

Needless to say, we would like to express our gratitude to all the authors for submitting their papers. They made it clear to us that there is a need for students to show their scientific products to the rest of the world and that we have to carry on publishing this annual edition. Also, we are extremely grateful to the reviewers for their elaborate comments to the papers. They make us feel confident that we are able to provide the reader with high quality student papers.

As you already might have noticed, the editorial board has welcomed a fourth member. Master's student Luís Tojo has reinforced the board and has been intensively involved in the complete process of this year's edition. Giving students a chance to play a role in the editorial process is completely in line with the aims of the journal. We strive to eventually have a faculty student journal in which only students conduct all activities. We hope that Luís has set the example and that more students are willing to become involved in the development of the third edition of the Maastricht Student Journal of Psychology and Neuroscience.

The editorial board

Anke Sambeth,  
Silke Conen,  
Luís Tojo,  
Tim Leufkens





# The Reliability of a Child as an Eyewitness in Court

## REVIEW

Whether eyewitness reports provided by children during a criminal court case are reliable, is frequently questioned. Factors that can influence the reliability of these reports are children's memory capacity, their susceptibility to suggestion, and the delay between a crime and providing an eyewitness statement. Eyewitness reports provided by children can be reliable given that this delay remains within a reasonable time frame, and that the presented questions are not suggestive. Additionally, eyewitness reports provided by older children are more reliable than those of younger children. A potential mechanism to increase the reliability is to use relevant cues or objects present at the time of the incident when the child is presenting evidence in court. Taking these factors into account in future criminal court cases with children as eyewitnesses will ensure the best possible reliability in children's statements, leading to an increased number of rightful convictions.

**Keywords:** children; eyewitness; reliability; memory; susceptibility

Ingeborg I.H.A. Close, Master student Psychology & Law  
Maastricht University, Maastricht, The Netherlands

---

iiha.close@student.maastrichtuniversity.nl

## INTRODUCTION

An eyewitness is an individual who has heard or seen anything regarding a criminal offense, and can therefore possess relevant information regarding this offense (De Rechtspraak, 2008). If such an individual is willing to serve as an eyewitness in court, he or she is likely to be able to prevent more crimes from occurring and help protect other individuals from becoming a victim of a crime (The Crown Prosecution

Service, n.d.). The number of children as eyewitnesses is ever-growing and therefore child eyewitnesses are more involved in the field of legal testimony (Bruck & Ceci, 1999; Flin, Boon, Knox, & Bull, 1992). Because of this greater involvement, it is frequently questioned whether children are able to serve as credible eyewitnesses during a criminal court case, especially in cases where the sole eyewitnesses to an offense are children (Flin et al., 1992). There are several factors that might influence the reliability of children's eyewitness reports. Firstly, children of all ages have the capability to give accurate reports when they are asked to freely recall a particular event. However, the younger the children are, the less detail they will eventually report (Pipe, 1996). Secondly, the delay between being a victim or witnessing a crime and providing an eyewitness statement can take up to six months (Flin et al., 1992). Since memory has the tendency to decline over time, the accuracy of the eyewitness' memory may decline as well (Law Commission, 1999). Lastly, children as well as adults appear to be suggestible, although younger children (5- to 8-year-olds) more so than older children (9- to 12-year-olds) and adults (Bruck & Ceci, 1999).

In the current paper, concern is raised regarding the accuracy of children's eyewitness reports, because of the above-mentioned factors. Therefore, in this paper it is tried to answer the question whether eyewitness statements provided by children are a reliable source to use in a criminal court case. When the factors affecting the reliability of children's eyewitness statements are known, specific methods to alleviate these factors can be designed, such as using suggestion free questions during children's eyewitness statements.

## RELIABILITY OF EYEWITNESS STATEMENTS

### **Memory capacity**

As mentioned previously, children are able to provide an accurate report when asked to freely recall a particular event, although younger children tend to report less detail than older children and adults (Pipe, 1996), which is problematic when serving as an eyewitness in court. Pipe and Wilson (1994) examined whether or not the memory capacity of children can be enhanced (i.e., providing more detail in their statements) by providing them with cues. The recruited children (6- and 10-year-olds) had an interaction with a magician, after which they were interviewed twice regarding this interaction (after 10 days and after 10 weeks). Additionally, the children were placed in one of four conditions: no cues (interview room not the same as magic show room), contextual cues (interview room the same as magic show room), relevant cues (items used by magician and contextual cues were present), and irrelevant cues (magic trick items similar to those used by the magician, and contextual and relevant cues were present).

It appeared that all children reported more accurate information after a short delay than after a long delay. However, younger children reported less accurate information than the older children. The relevant cues did facilitate free recall: all children reported more information when the relevant cues were present than when they were not. However, there was no difference in accuracy between the four conditions (Pipe & Wilson, 1994).

Gee and Pipe (1995) have performed a study which has replicated and extended the aforementioned results by conducting a similar study in 6- and 9-year-olds. This study showed that during free recall all children provided reports that were highly accurate. However, older children reported more correct information, but also made more errors than the younger children. After a short delay (10 days), all of the children reported more correct information than after a long delay (10 weeks). A prior interview increased the amount of information reported by the older children only after a long delay. It seemed that relevant objects attenuated the age differences that were present during prompted recall.

Although the results of both these studies have shown that children are able to provide accurate reports of experienced events, this accuracy can be negatively influenced by the way that children are questioned. Research has shown that the responses that children provide to open-ended questions (e.g., "Tell me what happened.") are more accurate than the responses they provide to specific questions (e.g., "Did you hurt your leg?") (Bruck & Ceci, 1999). Additionally, when children are presented with forced-choice questions (e.g., "Was it blue or red?"), they rarely indicate that they do not know the answer, which compromises the reliability of children's eyewitness reports (Bruck & Ceci, 1999). Also, repeated questioning can decrease the accuracy of children's responses to questions (Krähenbühl, Blades, & Eiser, 2009) as it can lead children to change their initial answer (Krähenbühl et al., 2009), perhaps because they assume incorrectly that their first response was incorrect (Memon & Vartoukian, 1996).

It appears that children are able to provide accurate reports of experienced events (Gee & Pipe, 1995; Pipe & Wilson, 1994), although they recall more correct information when they are presented with relevant cues (Pipe & Wilson, 1994) or prompts/objects (Gee & Pipe, 1995). Taken together, these findings indicate that when children are presented with relevant cues or objects that were present at the time of the incident, they might increase the reliability of the children's eyewitness reports. Additionally, to ensure the best possible reliability in children's statements, open-ended questions should be used during questioning.

### **Vulnerability to suggestion**

Another factor that can influence the reliability of children's eyewitness statements is vulnerability to suggestion, or the suggestibility effect (Ceci, Ross, & Toglia, 1987). When a certain memory trace or recollection of the original event becomes distorted or replaced after being exposed to erroneous post-event information, this is referred to as the suggestibility effect (Ceci et al., 1987). Ceci et al. (1987) designed experiments to investigate this effect in children. More precisely, they examined whether the memories of younger children are more vulnerable to misleading information than those of older children. All children were told a story after which they either received misleading information about the story or not. After an amount of time the children had to recall the story (Ceci et al., 1987). Results showed that the children most vulnerable toward the effects of misleading information appeared to be the youngest children (3- to 4-year-olds), whereas the other age groups (5- to 6-; 7- to 9-; and 10- to 12-year-olds) did not differ from each other. Furthermore, children that did not receive misleading information performed better than their

same age peers that did receive this information. This result has been replicated by two other studies performed by Ceci et al. (1987). The last study showed that children are susceptible to misleading information, regardless of whether this information is given by a child or an adult (Ceci et al., 1987). It has also been shown that children's suggestibility can be influenced by situational factors (Almerigogna, Ost, Bull, & Akehurst, 2007). Almerigogna et al. (2007) found that when children were being questioned by means of a non-supportive (e.g., serious behavior, closed body posture) instead of a supportive (e.g., friendly behavior, open body posture) interviewing style, they answered significantly more of the misleading questions incorrectly. This finding indicates that questioning children by means of a supportive instead of a non-supportive interviewing style could lead children to be more resistant to suggestions, and therefore keep children's suggestibility to a minimum (Almerigogna et al., 2007).

### **Delay between incident and statement**

An issue that is frequently questioned is whether children are able to recall accurate memories of a certain event a few months after this event has occurred, and this was studied by Flin et al. (1992). All included test subjects (5- to 6-year-olds, 9- to 10-year-olds and adults) observed an event after which they were either interviewed once (after a long delay) or twice (after a short and a long delay) regarding this event by means of cued recall (free recall of the event in combination with specific questions regarding what happened during the event) or enhanced recall (cued recall in combination with additional questions regarding contextual details of the event to enhance their memories). Results showed that the overall accuracy did not differ between the three age groups one day after the event. However, whereas the adults maintained their overall accuracy five months after the event, the overall accuracy of both children's age groups was significantly reduced and this reduction was largest for the younger children (Flin et al., 1992). Additionally, subjects who were interviewed by means of enhanced recall after day one had a significantly higher overall accuracy after five months, than those who had not been interviewed after day one. More recent research has shown that although children have a better verbal memory for a particular event after a short delay than after a long delay, they can have a relatively good verbal memory for an event that occurred six years ago (Jack, Simcock, & Hayne, 2012). The results indicate that although events can be verbally recalled after a long delay (Jack et al., 2012), the overall accuracy of children's eyewitness reports will be higher when witnesses are able to present their evidence within a short time frame after the incident has occurred (Flin et al., 1992). This time frame should be smaller for younger children, due to the greater loss in accuracy of their reports.

## DISCUSSION

Whether children are able to serve as credible eyewitnesses during a criminal court case is frequently questioned, and especially in those cases where the sole eyewitnesses to an offense are children (Flin et al., 1992). Studies regarding children's memory capacity have shown that children can provide accurate reports of events when they are asked to freely recall these events (Gee & Pipe, 1995; Pipe & Wilson, 1994), although older children provide more accurate information than younger ones. In the presence of relevant cues (Pipe & Wilson, 1994) or relevant prompts/objects (Gee & Pipe, 1995), children can recall more correct information, although younger children appear to be less accurate than older children in the presence of objects (Gee & Pipe, 1995). It also appeared that younger children are more vulnerable toward the effects of misleading information than older children (Ceci et al., 1987). This suggests that children, and especially younger children, are likely to agree when they are presented with questions that contain suggestions. However, it has also been shown that the suggestibility of children can be influenced by situational factors (e.g., interviewing style) during the questioning of children (Almerigogna et al., 2007). When using a supportive instead of a non-supportive interviewing style, children are likely to be more resistant to suggestions. Regarding the effect of a delay on memory, it has been shown that children are able to verbally recall a certain incident after a long delay (Jack et al., 2012). However, the reliability of the eyewitness reports provided by children is higher when the witnesses are able to present their evidence within a short time frame after the incident has occurred (Flin et al., 1992). This time frame should be smaller for younger children, due to the greater loss in accuracy of their reports.

Based on these findings, specific methods can be identified to facilitate the acquisition of more reliable eyewitness statements. Eyewitness reports provided by children can be reliable, provided that the questions presented to them are open-ended and do not contain suggestions, which children are likely to agree with. Also, while questioning children, the interviewer should adopt a supportive, instead of a non-supportive interviewing style, and the delay between the incident and providing a statement as an eyewitness should remain within a reasonable time frame. A potential mechanism to increase the reliability is to use relevant cues or objects that were present at the time of the incident when the child is presenting his or her evidence during a criminal court case.

As mentioned previously, the number of children as eyewitnesses is ever-growing (Bruck & Ceci, 1999; Flin et al., 1992). Therefore, ensuring the best possible reliability in children's statements will help lead to an increased number of rightful convictions. In future criminal court cases with children as eyewitnesses, the interviewer should adopt a supportive interviewing style and should avoid specific, forced-choice and repeated questions as well as the use of questions that contain suggestions. Also, the time frame in which children are summoned to present their evidence should be small, and children could be provided with relevant cues or objects to improve the accuracy of the eyewitness statements, and therefore make their statements as reliable as possible.

## REFERENCES

- Almerigogna, J., Ost, J., Bull, R., & Akehurst, L. (2007). A state of high anxiety: How non-supportive interviewers can increase the suggestibility of child witnesses. *Applied Cognitive Psychology*, 21, 963-974.
- Bruck, M., & Ceci, S.J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, 50, 419-439.
- Ceci, S.J., Ross, D.F., & Toglia, M.P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, 116, 38-49.
- De Rechtspraak. (2008). *Being a witness in a criminal trial*. Retrieved January 15, 2012, from [www.rijksoverheid.nl/bestanden/documenten-en-publicaties-brochures/2011/03/10/being-a-witness-in-a-criminal-trial/26br2008g005.pdf](http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties-brochures/2011/03/10/being-a-witness-in-a-criminal-trial/26br2008g005.pdf)
- Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, 83, 323-336.
- Gee, S., & Pipe, M.-E. (1995). Helping children to remember: The influence of object cues on children's accounts of a real event. *Developmental Psychology*, 31, 746-758.
- Jack, F., Simcock, G., & Hayne, H. (2012). Magic memories: Young children's verbal recall after a 6-year delay. *Child Development*, 83, 159-172.
- Krähenbühl, S., Blades, M., & Eiser, C. (2009). The effect of repeated questioning on children's accuracy and consistency in eyewitness testimony. *Legal and Criminological Psychology*, 14, 263-278.
- Law Commission. (1999). *Total recall? The reliability of witness testimony*. Retrieved January 15, 2012, from [http://www.lawcom.govt.nz/sites/default/files/-publications/1999/08/Publication\\_64\\_225\\_MP13.pdf](http://www.lawcom.govt.nz/sites/default/files/-publications/1999/08/Publication_64_225_MP13.pdf)
- Memon, A., & Vartoukian, R. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology*, 87, 403-415.
- Pipe, M.-E. (1996). Children's eyewitness memory. *New Zealand Journal of Psychology*, 25, 36-43.
- Pipe, M.-E., & Wilson, J.C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology*, 30, 515-525.
- The Crown Prosecution Service. (n.d.). *Why should I be a witness?* Retrieved January 15, 2012, from <http://www.cps.gov.uk/education/11-18/being-a-witness.html>

# The effect of shared leadership on team performance in international undergraduate students

## ORIGINAL PAPER

The current study examined how shared leadership influences a teams' performance among international undergraduate students. Shared leadership is a team property where two or more leaders are present in a team, sharing the lead and associated responsibilities. The study used a longitudinal design with five consecutive measurements during a six week period to inspect the temporal development and the various factors influencing shared leadership. The statistical analysis showed that shared leadership changed over time and both shared leadership and the personality trait conscientiousness predicted team performance. Correlation between motivation and team performance and an interaction effect between motivation and shared leadership could not be confirmed. Implications of the findings are discussed with reference to the current literature.

**Keywords:** shared leadership; team performance; conscientiousness; longitudinal study; international undergraduates

Kimani K. R. Michalke; Master Student Work & Organizational Psychology  
Maastricht University, Maastricht, the Netherlands

---

[kk.michalke@student.maastrichtuniversity.nl](mailto:kk.michalke@student.maastrichtuniversity.nl)

## INTRODUCTION

Organizations consider the development of leadership an important factor for their competitive advantage as it is crucial for team effectiveness (Hirst, Mann, Bain, Pirola-Merlo, & Richver, 2004; Sinclair, 1992; Zaccaro, Rittman, & Marks, 2001).



According to Pearce (2004), two particular shifts in the world of labor and economy fostered the query of the existing vertically-oriented leadership styles: First, intellectual capital utilized in organizations was growing to a degree that it can not be held by a single individual. Knowledge work (work that requires the extensive use of intellectual capital) is therefore increasingly divided among several individuals and becomes, as a result, team-based. Second, employees' attitudes changed; workers now desire a more meaningful impact on their working environment and progress. So the approach most discussed in recent literature is that of shared leadership (see e.g. Bligh, Pearce, & Kohles, 2006; Carson, Tesluk, & Marrone, 2007; Hoch, Pearce, & Welzel, 2010; Mehra, Smith, Dixon, & Robertson, 2006; Pearce, 2004; Pearce & Sims, 2002).

Even though the need for shared leadership was formulated in the 1930s already (Follet, 1924), mainstream literature ignored this approach. Instead, they distinguished between leaders and followers, and failed to recognize that leaders are not always appointed by higher authorities, but can hold other sources of power like superior knowledge or the degree to which they are favored by other members (Greenberg, 2011). In fact, high performing teams often have informal leaders (Neubert, 1999): Those with the most relevant knowledge or skill for a specific situation take the lead when the need arises. A combination of these upcoming views eventually led to the distributed leadership approach (Gronn, 2002) or shared leadership (Pearce & Sims, 2002). As defined by Carson et al. (2007), shared leadership emerges as a team property in which two or more team members engage in behavior that influences the direction and motivation of individual members and the team as a whole. Since the directing influence is distributed among members, responsibilities are shared likewise.

Although vertical leadership plays a role to form a team, lateral influences between team members are soon adapted and become the main source of team dynamics (Pearce & Sims, 2002). In accordance, Ensley, Hmieleski, and Pearce (2006) and Small and Rentsch (2010) found higher levels of shared leadership in mature teams than in newly assembled ones. However, Ensley et al. did not use a longitudinal design to confirm this assumption and Small and Rentsch only measured shared leadership at the start and the end of the study, whereas frequent measurements would provide more detailed insight into the development of shared leadership.

Regarding the effectiveness of shared leadership, recent studies mainly focus on the correlation between shared leadership and team performance. The general idea behind the concept of shared leadership is that the existence of several leading individuals in a team fosters participation and information-sharing, which in turn enhances performance (Mehra et al. 2006). However, results are inconsistent: Whereas some empirical studies found shared leadership to be a good predictor of team performance (see e.g. Pearce & Sims, 2002; Ensley et al., 2006; Carson et al., 2007), others only partly confirmed the model (see e.g. Hoch, 2007) or failed to report any significant correlations (see e.g. Mehra et al., 2006; Boies, Lvina, & Martens, 2010). Inferring from the above, shared leadership is an "approach under construction": Few influencing factors have been identified and tested empirically. One of the factors influencing shared leadership effectiveness might be the difference



in personality traits that are often associated with specific types of behavior. Some of these behaviors can be very conducive to team work, as stated by LePine and Van Dyne (2001). It was shown in their study that especially *Conscientiousness* is positively correlated to cooperative performance. This might foster behavior associated with shared leadership and, therefore, enhance team performance.

Another important factor is the common disadvantage of team work, as pointed out by Latané, Williams, and Harkins (1979): Whenever work load and responsibility are shared, there is a certain risk of free-riding (the tendency to be less productive in a group than when working individually) and social loafing (intentionally benefiting from group efforts without taking share in that effort). Teams containing one or more free riders are likely to be less effective and benefit less from shared leadership than teams consisting of exclusively motivated members.

To explore the temporal development of shared leadership, inspect its effectiveness, and empirically test the above made assumptions about influencing factors, the following hypotheses are proposed:

$H_1$ : Shared leadership increases over time.

$H_2$ : Shared leadership is positively related to team performance.

$H_3$ : Overall motivation of a team is positively related to team performance.

$H_4$ : Conscientiousness is positively related to team performance.

$H_5$ : Motivation positively moderates the relationship between shared leadership and team performance. More specifically: When motivation among team members is high, shared leadership and team performance are more positively related than when motivation is low.

## METHOD

### Procedure

Data for this study were collected at the University of Oslo from small groups of students. Participants were recruited from the undergraduate psychology course "Introduction to Organizational Psychology" (lasting one semester). Participation was voluntary, was not part of the academic curriculum, and was not rewarded financially or otherwise.

### *Formation of groups and group work*

In addition to weekly lectures and seminars, students were obliged to prepare and hand in written assignments consisting of case studies on the different course topics. Groups for these assignments were formed by the students in a self-organized manner. However, constellation of the groups had to be stable throughout the period of the study (no switching of members). Every group had to hand in a total of five assignments; one every week, for five consecutive weeks. Students would meet up, collect information, and distribute tasks autonomously. Workload was approximated to 8 hours per assignment (although it was not monitored systematically).

### *Weekly questionnaires*

Every week, after the respective assignment had to be handed in, students were asked to fill in the weekly questionnaire. Each questionnaire referred to the latest assignment respectively and contained the same questions for all five measurements. Participants filled in a total of five weekly questionnaires. They were informed about the procedure at the beginning of the study and were reminded to complete the questionnaire every week during seminars and via email. For every questionnaire, a due date was set at one week after the corresponding assignment had to be handed in to ensure comparability of the different measurement times.

### *Personal questionnaire*

Students were asked to fill in one additional questionnaire to gather information about characteristics that remain relatively stable over time, such as personality and gender. This personal questionnaire could be submitted any time from the beginning of the study until two weeks after the deadline of the last assignment.

The questionnaires (weekly and personal) were constructed with UniPark (psychology online survey software). All questionnaires were accessible online, meaning that students could choose locality and time of completion (e.g. using home computer, facilities of the university, or other instances). The questionnaires were administered in English.

## **Participants**

The sample consisted of 24 small international student groups, containing seventy eight subjects ( $N=78$ ). Overall response rate was 31% (personal questionnaire: 50%; weekly questionnaires: 27%). For the weekly questionnaires, 31% provided data on only one or two measurements; 69% provided data on three or more measurements. Three groups did not return any of the questionnaires and were excluded from the study. Group size was held relatively constant with a maximum number of four members per team because smaller groups are more effective than larger groups (Levine & Moreland, 1990). 65% of the participants were Norwegian; the other 35% were exchange students from countries other than Norway (including Germany, the Netherlands, Finland, France, Singapore, Spain, and Sweden). The mean age was 23.78 ( $SD= 4.16$ ), and 80% of the participants were female.

## **Measures**

At the start of every questionnaire, participants were informed about the aim of the study and were secured that their information would be treated confidentially.

### **Weekly questionnaires**

*Team performance* was assessed using questions by Hirst et al. (2004) and a five-point Likert scale (1=*disagree strongly*, 5=*agree strongly*). One of the questions used was: "The team has chosen the best available strategies for meeting project

goals.” Interitem correlations (expressed in Cronbach’s  $\alpha$ ) were calculated to inspect the internal consistency of items corresponding to the same concept (e.g. team performance). Coefficients are stated on the diagonal of Table 1. Cronbach’s  $\alpha$  was .79 at Time 1, .84 at Time 2, .85 at Time 3, .70 at Time 4, and .77 at Time 5. Measurements at Time 2 and 3 delivered very good Cronbach’s  $\alpha$  and can therefore compensate for slightly lower Cronbach’s  $\alpha$  at Time 4.

*Shared leadership* was assessed using questions by Hoch et al. (2010) and a 21-point Likert scale slide (1=*not at all*, 21=*to a great extent*). An example of one of the questions is: “My team members are driven by higher purposes or ideals.” Cronbach’s  $\alpha$  was .83 at Time 1, .84 at Time 2, .84 at Time 3, .82 at Time 4, and .91 at Time 5.

Group members rated team performance and the group’s leadership always in reference to one specific period of time (the period during which the group prepared their latest assignment) and in reference to the whole group rather than to individual members of the group (performance and leadership of the group during the completion of their latest assignment).

### *Personal questionnaire*

*Personality* was assessed using questions from the HEXACO-Personality Inventory-Revised (Lee & Ashton, n.d.) and a five-point Likert scale (1=*disagree strongly*, 5=*agree strongly*). Since the objective was to exclusively cover participants’ scores on the dimension conscientiousness, only questions referring to this specific personality trait were used. An example of one of the items is: “I always try to be accurate in my work, even at the expense of time.” A reversed example reads as: “When working, I sometimes have difficulties due to being disorganized.” Cronbach’s  $\alpha$  for personality was .84.

**Table 1.** *Descriptive Statistics and Standardized Correlation Coefficients*

<i>Variables</i>	Mean	SD	1	2	3	4
1 Team performance <sup>a</sup>	4.02	.49	<b>.79<sup>b</sup></b>			
2 Shared leadership <sup>a</sup>	3.55	.66	.65***	<b>.85<sup>b</sup></b>		
3 Motivation	3.76	.55	.16	.17	<b>.76</b>	
4 Personality	3.68	.57	.10	-.19	.42**	<b>.84</b>

*Note:* Interitem correlations are presented on the diagonal.

SD = Standard Deviation

<sup>a</sup> Level 1-variables (repeated measures) have been aggregated to Level 1 (individual).

<sup>b</sup> For level 1-variables, interitem correlations are stated as the average Cronbach’s  $\alpha$  over all 5 measurement times. Cronbach’s  $\alpha$ s for the individual measurement times are .79, .84, .85, .70, .77 for team performance and .83, .84, .84, .82, .91 for shared leadership.

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

*Motivation* was assessed using adjusted questions by Price (1997) and a five-point Likert scale (1=*disagree strongly*, 5=*agree strongly*). Price's original inventory refers to a work situation and individual tasks, whereas the present study was conducted in an academic context and with reference to overall group work. Questions were therefore adapted to better suit the aim of this study: "job" was replaced by "group work" and "my" was replaced by "our". Questions as "I feel a sense of personal satisfaction when I do this *job* well." and "I feel unhappy when *my* work is not up to my usual standard." became "I feel a sense of personal satisfaction when I do this *group work* well." and "I feel unhappy when *our* work is not up to my usual standard." Cronbach's  $\alpha$  for motivation was .76.

## Analysis

The present study contained two levels of analysis: repeated measurements over time (Level 1) and between individuals (Level 2). Level 1-variables are shared leadership and team performance. Level 2-variables are personality, motivation, and the control variables age, gender, and nationality. Missing values were replaced using missing value analysis and a subsequent application of the EM method (SPSS 19).

### Pre-analysis

For the pre-analysis, Level 1-variables have been aggregated to Level 2. Independent samples t-tests were carried out for age, personality, and intrinsic motivation, and a  $\chi^2$ -test for gender to inspect possible differences between students who provided data on only one or two measurements versus students who provided data on three or more measurements. Results showed that students who filled in just one or two weekly questionnaires did not differ from students who filled in three or more questionnaires. Independent samples t-tests were carried out to examine possible differences between Norwegian and non-Norwegian participants with regard to personality, motivation, shared leadership, and team performance. Norwegian participants did not differ from non-Norwegian participants on these variables.

Descriptive statistics and correlation coefficients have been calculated for team performance, shared leadership, motivation, and personality (see Table 1). Shared leadership was significantly related to team performance and motivation was significantly related to personality. Results of the Multilevel Analysis stated below confirm the general observations.

### Multilevel Analyses

To inspect the temporal development of shared leadership, a mixed model with a random intercept for the predictor variable time and shared leadership as the dependent variable was run. Regarding Multilevel analysis, a mixed model for repeated measures with team performance as the dependent variable was run. A model with only shared leadership as independent variable (corresponding to Model 3 of the below stated) was used to compare different error covariance structures. The following structures were tested: diagonal, autoregressive (first order), heterogeneous autoregressive, and heterogeneous compound symmetric. The goodness-of-fit statistic Akaike's Information Criterion (AIC) was consulted

to compare the performance of the applied covariance structures. Autoregressive structure best fitted the data of this study.

Consequently, multilevel analysis was conducted using the autoregressive error covariance structure. First, an unconditional model (Model 1) without predictors was run to inspect the significance of within-group variance (random residual estimate) and between-group variance (random intercept estimate), as is suggested by Hox (2002). The different predictor variables were then added stepwise to explore their relevance in explaining variation in the dependent variable team performance.

## Results

Results of the mixed model analysis of the temporal development of shared leadership can be seen in Table 2. The estimate for the predictor variable time is .07 and is significant at  $p < .05$ , indicating that for every one consecutive measurement the average shared leadership value increases by .07.

**Table 2** *Results of Mixed Model Analysis: Temporal Development of Shared Leadership*

<i>Variable</i>	<i>Est.</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>F</i>	<i>Sig.</i>
Time	.07	.03	31.85	2.23	4.98	.03

*Note:* Every unit increase in the predictor variable results in a change of the value of the dependent variable corresponding to the coefficient of the predictor estimate.

Est.= Predictor Estimate, SE= Standard Error, df= Degrees of Freedom, t= t-test value,

F= ANOVA test value, Sig.= Significance

Results of the different models of the multilevel analysis can be seen in Table 3. Akaike's Information Criterion was used to determine which Model fitted best the data (lower coefficient indicates better fit of the model). With a value of 121.23, the model that included time, shared leadership, motivation, and personality (Model 5) best fits the data. The predictor estimates for shared leadership and personality in this model appear to positively predict team performance. The estimate for shared leadership is .57 ( $p < .001$ ), indicating that an increase in participants' shared leadership value of 1, results in an increase of team performance equal to .57. The estimate for personality is .23 and is significant at  $p < .01$ : Therefore, an increase in the personality trait conscientiousness (which was measured and coded in this study as personality) results in an increase in team performance of .23. Introducing the variable shared leadership to the multilevel analysis (Model 3) leads to a decrease in total variance from .30 to .18. This is a decrease of 40% compared to the previous model. It can therefore be assumed that 40% of the total variance can be explained by shared leadership. When adding the predictor personality to the analysis (Model 5), total variance decreases from .18 to .16. This is a decrease in total variance of 11%, indicating that the Level 2 predictor personality explains 11% of the total variance. Time and motivation do not improve significantly the multilevel model and adding these variables does not explain more variance of the dependent variable (Model 2 & 4). A moderating effect of motivation (interaction) was not found (Model 6).

Table 3 Results of Multilevel Analysis: Predictors of Team Performance

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<i>Variables</i>												
Intercept <sup>a</sup>	3.97***	.09	3.87***	.12	2.15***	.23	2.23***	.38	1.5**	.44	3.80*	1.71
time			.04	.03	-.01	.03	-.01	.03	-.01	.03	-.00	.03
SL					.53***	.06	.54***	.07	.57***	.03	-.06	.45
Motivation							-.03	.10	-.09	.10	-.66	.42
Personality									.23**	.08	.22**	.08
SL× motivation											.16	.11
<i>Variance</i>												
Total variance		.30		.30		.18		.18		.16		.20
Diff. variance <sup>b</sup>		-		0 %		40%		0%		11%		- 25%
<i>Fit index</i>												
AIC <sup>c</sup>		160.23		161.10		124.98		126.92		121.23		121.36

Note: Every unit increase in the predictor variable results in a change of the value of the dependent variable corresponding to the coefficient of the predictor estimate. Est.= Predictor Estimate, SE= Standard Error, SL= Shared Leadership, AIC= Akaike's Information Criterion

<sup>a</sup> Predicted value of the dependent variable when all predictors are set to zero. <sup>b</sup> Reduction in total variance compared to previous model. <sup>c</sup> Lower coefficient indicates better fit of data.

\* $p<.05$ ; \*\* $p<.01$ ; \*\*\* $p<.001$

## DISCUSSION

The goal of the present study was to identify various factors that influence the emergence and effectiveness of shared leadership. Results show that shared leadership increases over time and that higher levels of shared leadership are associated with higher team performance. Furthermore, the personality trait conscientiousness is positively correlated with team performance. A hypothesized effect of motivation on team performance could not be confirmed. Neither did motivation moderate the relationship between shared leadership and performance.

A main contribution of the present study was the longitudinal study design aiming to track the temporal development of shared leadership. The steady increment of shared leadership accommodates an assumption made by Ensley et al. (2006): The relative importance of vertical versus shared leadership depends upon the stage of development of the team. Whereas established teams rely on shared leadership, young teams need a more vertical leader. As teams in this study were formed at the starting point of the investigation and then matured over the course of the study, shared leadership became increasingly important and more persistent. As noted by Ensley et al. (2006), in case of complications or external changes, the degree of shared leadership drops and a vertical leader is needed again. As a result, leadership might change over the lifespan of a team in a wavelike manner.

Even though the effectiveness of shared leadership has only recently been tested empirically and the body of research is still growing, shared leadership seems to comply with the rising demand for flexibility and efficiency in organizations and the aspiration among employees to partake in decisions. Consequently, shared leadership has been introduced as the new, more suitable leadership approach (Hoch et al., 2010). However, as suggested by Fiedler's contingency model of leadership effectiveness, the best-fitting leadership style depends on the situational context (Greenberg, 2011; Vroom & Jago, 2007). According to Fiedler, a certain leadership style might be effective in some situations but not in others. In surgical teams or combat troops for instance, information has to be passed on without delay, so that action can be initiated rapidly. These teams will likely benefit from a more hierarchical structure, while the need for innovation is a minor concern. The theoretical model of shared leadership suggests that it allows a more complete usage of the available intellectual resources and fosters participation, which in turn enhances team performance (Mehra et al., 2006). This is especially true when the required intellectual resources are extensive and rapid processing is of little importance. Under these circumstances, shared leadership is more suitable.

Since the two general leadership views – shared versus vertical leadership – cannot be distinguished in a real life setting in such a clear cut manner as they can be in theory, Gronn (2002) stated that they are best described as opposing positions on a continuum. Identifying factors that predict effectiveness for one or the other extreme can help approximate the best fitting style on the continuum. The present study shows that undergraduate work teams benefit from a more shared leadership, thereby amplifying the applicability of shared leadership to the academic environment.

A drawback of shared leadership research is that most published studies have been conducted in North America and only very few have been conducted in



other parts of the world (for a German sample study see Hoch et al. 2010). Cultural differences, however, might affect the emergence, development, and effectiveness of shared leadership. As can be inferred from Hofstede's cultural dimensions theory (Hofstede & Hofstede, 2005), cultures differ in the degree to which members accept the unequal distribution of power in organizations and institutions. This difference in acceptance of disparity or "power distance" indicates that shared leadership might not be equally applicable in all cultures. It is therefore important to identify cultures that likely benefit from this leadership style. The present study adds to recent findings by replicating results regarding its effectiveness in an international population.

Besides environmental factors, there are notable individual characteristics that evoke differences in shared leadership and affect team effectiveness. One of these characteristics, as identified in the current study, is the personality trait conscientiousness: Higher levels of conscientiousness correlate with higher team performance. This finding is consistent with recent literature as conscientiousness is usually associated with cooperative behavior (LePine & Van Dyne, 2001), which enables shared leadership and therefore enhances performance.

Another crucial characteristic of a team is the motivation of their members: Poorly motivated teams are less productive than motivated teams due to social loafing (the tendency to be less productive in a group than when working individually) and free riding (intentionally benefiting from group efforts without taking share in that effort). A hypothesized positive correlation between motivation and team performance, however, was not confirmed in this study, implying that poorly motivated teams perform as well as highly motivated teams. This finding is highly counterintuitive since empirical work has shown that motivation is an important predictor of performance (Latané et al., 1979). A possible statistical explanation for the non-significant motivation effect might be that the shared leadership variable already covers most of the variance explained by motivation. A follow-up analysis produced a positive trend for motivation to correlate with team performance that was significant at  $p < .10$ .

As pointed out by Hoch et al. (2010), the correlation between shared leadership and team performance found in recent studies varies somewhat across studies. Hence, it has consequently been argued that future research should zoom in on negative moderator effects in the leadership-performance relationship (Avolio, Walumbwa, & Weber, 2009; Mohammed & Nadkarni, 2011; Pearce & Conger, 2003). As seen in other empirical work, motivation sometimes acts as a mediator between certain predictor variables and performance (Gagné & Deci, 2005), and sometimes as a moderator of the predictor variables (Dysvik & Kuvaas, 2011). Highly motivated teams might display the actual effect of shared leadership on team performance, whereas poor motivation might interfere with the shared leadership-performance relationship. An interference or moderation effect, however, was not significant in this study, indicating that shared leadership predicts team performance independent of the team's motivation.

In conclusion, following from the results of this study, strategic shared leadership training should not only be offered to teams in organizations and firms, as suggested by Carson et al. (2007), Ensley et al. (2006), and Hoch et al. (2010), but also to students in an academic context.



### **Future research**

Several suggestions for future research are readily identifiable from the results of this study. First, the present study used a longitudinal study design to inspect the development of shared leadership in teams that were newly assembled and then persisted for one semester. Although the present study adds to the available knowledge about shared leadership as most studies did not focus on its development over time, future studies should examine shared leadership with an even more longitudinal research. By doing so, researchers would be able to detect possible factors involved in a change of leadership style as a result of evolving problems, changing environmental factors, or switching of team members.

Second, even though the present study extended the applicability of shared leadership to non-U.S. populations, further intercultural comparison is required to zoom in on differences between collectivistic and individualistic oriented societies. A research topic for future studies could be the possible moderator effect of culture for the relationship between shared leadership and team performance.

Third, the present study only inspected the impact of conscientiousness on team performance, but not that of other personality traits. Future research should explore the effect of different personalities on team performance and check for possible mediator effects of personality.

### **Limitations**

This research was not without limitations. First, the sample of this study was drawn from attendees of an undergraduate organizational psychology course. On one hand, as mentioned earlier, the use of such a sample extends the applicability of shared leadership to an academic context (most studies so far have only reported positive effects of shared leadership in full-time working environments). On the other hand, the specificity of the sample reduces the generalizability of the findings.

Second, a high percentage of subjects in this study were females (80%) and individualistic cultures were more represented in the sample than collectivistic cultures (for a classification of cultures see Hofstede & Hofstede, 2005). Differences between men and women regarding leadership style and performance are usually of minor concern as pointed out by Burke and Attridge (2011) and Lally (2008). Applicability of shared leadership might, however, differ depending on the cultural context. Results can therefore not be generalized to the same extent to collectivistic and individualistic cultures.

Third, questionnaires were administered in English. Even though all participants had a good understanding of English (language of instruction within the academic setting was English), the majority was not native speakers. Differences in item comprehension due to completion of questionnaires in a foreign language might be present.

Finally, no conclusion can be drawn about the causality of effects found in this study: Shared leadership might be the cause of high performance among undergraduate students. It might also be the case that high performing teams

are simply more likely to engage in shared leadership, or that high performance just happens to co-occur with shared leadership as a result of a third, unidentified variable.

## Conclusion

In conclusion, this study shed light on a new and unexplored approach of leadership. The findings of the study add to the growing body of evidence that shared leadership improves team effectiveness. They also contribute to the general understanding of the shared leadership approach by broadening its generalizability to non-U.S. countries and by specifying circumstances in which the approach might be applicable (undergraduate students in an academic context). Managers and academics should therefore take into account and invest in shared leadership by offering training to employees and students.

## Acknowledgements

I would like to thank Dr. Sabine Raeder for her assistance with data analysis and revision. Robert van Doorn and an anonymous reviewer also provided useful comments and suggestions on previous versions of this article.

## REFERENCES

- Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology*, 60, 421-449.
- Bligh, M. C., Pearce, C. L., & Kohles, J. C. (2006). The importance of self- and shared leadership in team based knowledge work: A meso-level model of leadership dynamics. *Journal of Managerial Psychology*, 21(4), 296-318.
- Boies, K., Lvina, E., & Martens, M. L. (2010). Shared leadership and team performance in a business strategy simulation. *Journal of Personnel Psychology*, 9(4), 195-202.
- Burke, J. M., Attridge, M. (2011). Pathways to career and leadership success: Part 2 - striking gender similarities among \$100k professionals. *Journal of Workplace Behavioral Health*, 26, 207-239.
- Carson, J. B., Tesluk, P. E., & Marrone, J. A. (2007). Shared leadership in teams: An investigation of antecedent conditions and performance. *Academy of Management Journal*, 50(5), 1217-1234.
- Dysvik, A., & Kuvaas, B. (2011). Intrinsic motivation as a moderator on the relationship between perceived job autonomy and work performance. *European Journal of Work and Organizational Psychology*, 20(3), 367-387.
- Ensley, M. D., Hmieleski, K., M., & Pearce, C. L. (2006). The importance of vertical and shared leadership within new venture top management teams: Implications for the performance of startups. *The Leadership Quarterly*, 17, 217-231.
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26, 331-362.
- Greenberg, J. (2011). Behavior in organizations. Global Edition: Pearson.
- Gronn, P. (2002). Distributed leadership as a unit of analysis. *The Leadership Quarterly*, 13, 423-451.
- Hirst, G., Mann, L., Bain, P., Pirola-Merlo, A., & Richver, A. (2004). Learning to lead: The development and testing of a model of leadership learning. *The Leadership Quarterly*, 15, 311-327.

- Hoch, J. E. (2007). *Verteilte Führung in virtuellen Teams. Zum Einfluss struktureller, interaktionaler und teambasierter Führungstechniken auf den Teamerfolg*. Unpublished doctoral dissertation, University of Kiel, Kiel. Retrieved from: [http://eldiss.uni-kiel.de/macau/servlets/MCRFileNodeServlet/dissertation\\_derivate\\_00002122/dissertationsschrift\\_julia\\_hoch.pdf?hosts](http://eldiss.uni-kiel.de/macau/servlets/MCRFileNodeServlet/dissertation_derivate_00002122/dissertationsschrift_julia_hoch.pdf?hosts)
- Hoch, J. E., Pearce, C. L., & Welzel, L. (2010). Is the most effective team leadership shared? The impact of shared leadership, age diversity, and coordination on team performance. *Journal of Personnel Psychology*, 9(3), 105-116.
- Hofstede, G., & Hofstede, G. J. (2005). *Cultures and organizations: Software of the mind*. NY: McGraw Hill.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Lally, P. (2008). *The relationship between gender and leadership style in the Massachusetts high school principalship*. Dissertation abstract. Retrieved from: ProQuest Dissertations & Theses.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6), 822-832.
- Lee, K. & Ashton, M. C. (n.d.). Self-Report Form HEXACO-60. In *The HEXACO Personality Inventory-Revised*. Retrieved September, 2011, from <http://hexaco.org/>
- LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential effects of Big-Five personality characteristics and general cognitive ability. *Journal of Applied Psychology*, 86, 326-336.
- Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 41, pp. 585-634). Palo Alto, CA: Annual Reviews.
- Mehra, A., Smith, B. R., Dixon, A. L., & Robertson, B. (2006). Distributed leadership in teams: The network of leadership perceptions and team performance. *The Leadership Quarterly* 17, 232-245.
- Mohammed, S., & Nadkarni, S. (2011). Temporal diversity and team performance: The moderating role of team temporal leadership. *Academy of Management Journal*, 54(3), 489-508.
- Neubert, M. J. (1999). Too much of a good thing or the more the merrier? Exploring the dispersion and gender composition of informal leadership in manufacturing teams. *Small Group Research*, 30(5), 635-646.
- Pearce, C. L. (2004). The future of leadership: Combining vertical and shared leadership to transform knowledge work. *Academy of Management Executive*, 18(1), 47-57.
- Pearce, C. L., & Conger, J. A. (2003). *Shared Leadership: Reframing the Hows and Whys of Leadership*. Thousand Oaks, CA: Sage.
- Pearce, C. L., & Sims Jr., H. P. (2002). Vertical versus shared leadership as predictors of the effectiveness of change management teams: An examination of aversive, directive, transactional, transformational, and empowering leader behaviors. *Group Dynamics: Theory, Research, and Practice*, 6(2), 172-197.
- Price, J., L. (1997). *Handbook of organizational measurement* (pp. 411- 421). Iowa, USA: MCB UP Ltd.
- Sinclair, A. L. (1992). The tyranny of a team ideology. *Organization Studies*, 13, 611-626.
- Small, E. E., & Rentsch, J. R. (2010). Shared leadership in teams. A matter of distribution. *Journal of Personnel Psychology*, 9(4), 203-211.
- Vroom, V. H., & Jago, A. G. (2007). The role of the situation in leadership. *American Psychologist*, 62(1), 17-24.
- Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *Leadership Quarterly*, 12, 451-483.

# The Effects of Mindfulness Training on Children's Attention Skills

## ORIGINAL PAPER

**Introduction:** Mindfulness Trainings (MT) have been shown to enhance participants' attention skills by improving attention networks as well as supporting emotion regulation which may otherwise lead to distractions or biased attention. While research in adult populations shows promising results, only a few studies have examined the effects of MT in children. This investigation aims to address the gap in research by assessing the effects of mindfulness training for children at two schools (school 1: N=24, aged 9-12; school 2: N=63, age 8-11).

**Methods:** Children were assessed on the Attention Network Task - Interference (ANT-I; N=91) and on an Attention Control Task (N=91). Furthermore, parents filled in questionnaires rating children's executive functions. Measurements were assessed before and immediately after the 5-6-week training. School 1 MT group (N=12) was compared to an active reading control group (N=12), and school 2 MT group (N=24) was compared to an active reading control group (N=22) as well as a passive control group (N=21).

**Results:** Performance on the ANT-I and on the Attention Control Task did not show a significant main effect for training group. In order to explore whether children's emotional regulation improved attention, children's performance was grouped in high and low emotional control group, which did not differ significantly pre- and post-test on main effects on ANT-I.

**Conclusion:** While mindfulness trainings hold many promises in adult populations, this study was unable to show that mindfulness training in children enhances their attentional control. This may be due to the young age of the children or the relatively short period of MT. Further studies need to be conducted in order for sufficient conclusions to be drawn about the effects of mindfulness training with children.

**Keywords:** mindfulness, emotions, attention

Alice Wellum; master student Developmental Cognitive Neuroscience  
Maastricht University, Maastricht, the Netherlands

---

a.wellum@student.maastrichtuniversity.nl

## INTRODUCTION

A growing body of research shows longitudinal relations between children's executive functions (such as attention, inhibition, working memory and planning), social-emotional development and their academic performance (Blair & Razza, 2007; Hughes, Cutting, & Dunn, 2001; Riggs, Blair, & Greenberg, 2003).. The need to support children especially in their attentional capacities is best reflected in the growing body of research investigating how children's attention skills can be enhanced. While several trainings, such as visuo-spatial computer games (Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2008) and the Attention Process Training (Kerns, Esso, & Thompson, 1999) have been developed, many trainings do not take the mediating role of emotions on attention into account. For example, anxiety has been shown to affect attentional biases to threatening stimuli in children, sometimes at the cost of attention for other potentially relevant information (Ehrenreich & Gross, 2002; Taghavi, Neshat-Doost, Moradi, Yule, & Dalgleish, 1999). It is therefore not enough to only train attention processes. In order to enable children to fully access their attention skills they must also be supported in their emotion regulation. A training which addresses both emotion regulation *and* attention networks is mindfulness based training (MT). MT has been shown to improve attention networks in adult populations (Jha, et al., 2007), and to reduce negative emotional affect (Tang et al., 2007; Arch & Craske, 2006).

This study examined the effects of MT on children's attention skills using two measures of attention: the Attention Network Test (ANT) and an Attentional Control Task (Alberts, Martijn, Greb, Merckelbach & de Vries, 2007), which are described below. Furthermore, the link between emotional control and attention was examined by assessing whether children with low emotional control (as rated by parent-rated questionnaires) benefit more from MT than children with high emotional control. Before further explaining the design and tasks used in the present study, Posner's attention network model, on which the tests used in the current study are based, will be explained. This will be followed by a discussion on how emotional states can influence attention, which is investigated in the current study by examining a possible interaction between emotional control and MT. Subsequently, the design is outlined in detail. Finally, the results are discussed.

**Attention: A Tripartite Network Model and the Effects of Emotions on Attention**

According to Posner's attention network theory (Posner & Boies, 1971; Posner & Petersen, 1990), three different attention networks can be distinguished: the executive control network, the orienting network and the alerting network. While the executive control network is active when a person faces a situation that involves overcoming habitual actions and giving novel responses (Posner & Petersen, 1990), the orienting network, on the other hand, selects where to attend to in the environment (Posner & Petersen, 1990). Finally, the alerting network creates an internal state of high sensitivity to incoming stimuli (Posner, 2008).

In children, these three networks appear to mature at different ages. Trick and Enns (1998) report some very early development of alerting before the age of six with little improvements after this, while others such as Rueda and colleagues (2004) report that the alerting network seems to be mature by the age of four years already. In contrast, while simple orienting such as moving the eyes in response to a flash matures early (Baijal, Jha, Kiyonaga, Singh, & Srinivasan, 2011), more complex orienting processes such as the ability to select relevant information during distraction do not reach adult levels until the age of nine years (Huang-Pollock, Mikamo, Pfiffner, & McBurnett, 2007). Finally, the third type, executive control as measured by the ANT-I is thought to only be fully mature at age twenty-five (Rueda et al., 2004).

The efficiency of these individual networks is assessed during the Attention Network Test (see Figure 1). In this test, the executive control network is assessed via a 'flanker' task. A target (arrow or fish) is presented at every trial, with flankers that can either point in the target's direction (congruent condition) or the opposite direction (incongruent condition). Participants indicate by pressing one of two keyboard keys which direction the target is pointing at. In general, the reaction times in the congruent condition will be faster than those in the incongruent condition, since the incongruent condition will invoke some interference. The flanker task is preceded by a spatial cueing task, which assesses the efficiency of the orienting network. Here, an asterisk can appear in the position of the target (valid cue), in a different location (invalid cue), or it not will not appear at all (neutral cue). Thus, a valid cue will lead the subject to orient attention to the target location so that a very fast response can be given. An invalid cue will draw attention away from the target location so that disengagement of attention is required first before attention can be shifted to the target location. This disengaging takes time and the difference in reaction time (RT) between invalid and valid trials reflects the time needed to disengage and shift attention. The smaller the difference in RT between invalid and valid conditions, the more efficient one is in disengaging and orienting (shifting) attention across space. Finally, the efficiency of the alerting network can be assessed during trials with an audio cue that prepares the participant to respond to a visual cue at an unknown location (Fan et al., 2002).

It is important to note that these three networks can also be influenced by the emotional state of a person (Vasey, Daleiden, Williams, & Brown, 1995; Vasey, El-Hag, & Daleiden, 1996; Ehrenreich & Gross, 2002). In Posner's model of selective attention, Posner compares attention to a spotlight which illuminates certain locations at the expense of other, non-attended spatial locations (Posner,



1978; Treisman, 1969). Emotions can sometimes influence where this spotlight falls, allocating attention towards salient stimuli that are relevant to a person's internal state (Crick & Dodge, 1994) and guiding subsequent judgements and behaviours (Fisk & Taylor, 1991; Pashler, Johnston, & Ruthruff, 2001). For example, in an amended Stroop-task with spider-phobic children aged six and seven years, children demonstrated slower reaction-times for colour naming of spider-relevant words, while control children displayed no such bias (Martin, Horder & Jones, 1992). Additionally, non-clinical high-anxious children displayed a colour-naming bias associated with the presentation of general threatening stimuli, while low-anxious children demonstrated interference in colour naming of threatening words when an immediate stressor (e.g. an injection) was presented (Kindt, Brosschot, and Everaerd, 1997). Given the influence of emotions on attention, it follows that to effectively train children's attention one must also train their emotional regulation. MTs are one of the few trainings which do this. In the following section, the effects of MT on emotions and attention will be explored.

### **The Effects of Mindfulness on the Attention Networks and Emotional Regulation**

Mindfulness can be defined as “the awareness that emerges through paying attention on purpose, in the present moment, and nonjudgmentally to the unfolding of experience moment by moment” (Kabat-Zinn, 1994, p.144). Most mindfulness trainings focus on learning two kinds of meditation. The first one, called “concentrative” or Focused Attention (FA) meditation, involves training to focus attention on one single aspect such as the breath or a mantra. The second one is called “receptive” or Open Monitoring (OM) meditation and involves the attention being kept in the present moment of experience without limiting or directing it in any way. Extraneous stimuli are considered distractors in FA meditation, while in OM meditation attention is open towards all stimuli experiences (Jha et al., 2007).

Several studies suggest that FA meditation trains the executive control network and the orienting network, since it relies on the exclusion of unwanted stimuli and (re-)focusing on the desired object of control. In adult populations, case-control studies have shown that experienced practitioners of FA meditation demonstrated better executive control scores on the ANT-I than meditation-naïve participants did. Furthermore, eight weeks participating in a Mindfulness Based Stress Reduction (MBSR) training improved participants' orienting scores compared to those of an inactive control group (Jha et al., 2007). Following a brief MT intervention (twenty minutes a day over five days of training), participants significantly improved their executive control scores on the ANT-I compared to a control group (Tang et al., 2007).

Whereas FA meditation improves orienting and control networks, OM meditation appears to improve the alerting network. Experienced adult meditators participating in an OM mindfulness training improved their alerting scores on the ANT-I relative to meditation-naïve participants and participants who had taken part in an eight-week MBSR training. Furthermore, for the experienced meditator group alerting scores were correlated with prior meditation experience, with greater meditation experience correlating with reduced alerting scores (Jha et al., 2007).

Support for the interaction between emotion and attention can also be found

in various studies. Tang et al. (2007) found that after adults were given MT there were improvements in positive emotions and reduction in negative emotions as measured by the Profile of Mood States (POMS) questionnaire, indicating that participants gained more emotional self-regulation and control. A further study found that 15 minutes MT sessions could moderate emotional responses to affectively valenced picture slides compared to two active control groups who were either instructed to think about worrying thoughts or let their attention wander (Arch & Craske, 2006).

While the number of MT studies in children is still very limited, some studies have been done (see Table 1) and results will be reviewed shortly. Napoli and colleagues (2005) reported that MT improved selective attention and reduced anxiety in 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> grade children who participated in MT by the Attention Academy, compared to children from a reading-list control group. In clinical populations, MT has also shown to be effective for children with reading difficulties. Parents reported significantly reduced attention problems compared with parents of a waiting list control group (Semple et al., 2010). Furthermore, adolescents with Attention Deficit Hyperactivity Disorder (ADHD) reported reduction in self-reported ADHD symptoms, and showed improvements in measures of executive attention as measured by the ANT-I, a Stroop task and a Trail-making test (Zylowska et al., 2008). Finally, adolescents with ADHD, Oppositional Defiant Disorder (ODD)/Conduct Disorder (CD) and Autistic Spectrum Disorder (ASD) showed significant improvements in sustained attention on the D2 test of attention (Boegels, Hoogstad, van Dun, Schutter, & Restifo, 2008). However, no control groups were used for the latter studies and the findings should be viewed with caution. Finally, a case-report study reported improved alerting and executive attention rates on the ANT-I in children who had FA meditation experience of at least two years, when compared to meditation naïve children (Bajjal, Jha, Kiyaga, Singh, & Srinivasan, 2011).

While many of the above studies seem to indicate that MT has positive effects on children's attention and emotion, a note of caution is advised when interpreting these results. Firstly, as can be seen in Table 1, there is a wide discrepancy in methodology and study design, so a comparison between studies should be cautionary at the least. Secondly, most mindfulness studies involving children have focused on clinical populations; few have used typically developing populations. Thirdly, some studies suffer from methodological problems such as a lack of control groups, small sample sizes, and utilizing self-reports instead of performance tests or other more objective measures. Fourth and finally, there are vast differences in training programs that differ in content, length and duration of sessions. The present study was created in order to address some of the above shortcomings.

### **The Present Study**

The present study will focus on the effects of MT on children's attentional capacities. Passive and active control groups will also be included.

1. The first hypothesis of this study states that compared to the active and the passive control groups, MT will lead to:
  - a) Improved attentional performance in the ANT-I (indicated by lower RTs) and
  - b) Larger increase in correctly solved sums on the Attentional Control Task.



Table 1: Overview of Studies Examining the Effects of MT on Attention in Children

Author(s)	Training study	Children's ages/ clinical diagnosis	Type of MT	Attentional Results
Bajjal, Jha, Kivanaga, Singh & Srinivasan (2011)	No training study	13-15 Non-clinical population	Concentrative meditation training	Alerting and executive attention, but not orienting, was found to differ between the CMT and the control group as measure by the ANT. Executive attention demonstrates age-related improvements.
Semple, Lee, Rosa & Miller (2010)	Training study	9-13 reading difficulties	MBCT-C	Children in mindfulness group showed significant reductions in attention problems on the Child Behaviour Checklist as rated by parents compared to controls.
Boegels, Hoogstad, van Dun, Schutter and Restifo (2008)	Training study	11-18 ADHD, ODD/CD and ASD	Adapted MBCT programme	Significant improvement on sustained attention was found as measured by the D2 test of attention
Zylowska, Ackerman, Yang, Futrell, Horton, Hale, Pataki & Smalley (2008)	Training study	8 adolescents & 15 adults with ADHD	MAPS for ADHD	Self-reported reduction in ADHD symptoms, improvements on measures of attentional conflict on ANT.
Napoli, Krech & Holley, 2005	Training Study	First, second and third-grade children non-clinical population		Self-reported reductions in anxiety as measured by the Test Anxiety Scale, and increase in selective attention as measured by two visual attention measures.

Furthermore, given that early meditation training seems to affect the executive control network and orienting network in particular,

2. The second hypothesis states children's mindfulness scores will have improved more in the executive control network and orienting network than in the alerting network.

Finally, due to the above reviewed links between emotion regulation and attention, as well as the fact that MT has often been reported to positively affect emotion regulation, we also investigated whether pre-existing differences in emotion control might have influenced training effects on attention measures.

3. The third hypothesis therefore states that children who display low emotional control at pre-training will have a larger increase in attention performance on the ANT-I after following the MT than children with high emotional control who follow the MT.

## METHOD

### Participants

The sample consisted of 29 children in fifth and sixth grades with a mean age of 10.8 years ( $SD=0.85$ ), recruited from a primary school in Roermond (School 1), as well as 83 children recruited from a primary school in Swalmen (School 2) who had a mean age of 9.4 year ( $SD=0.66$ ). Parents and school were briefed and signed a consent form. The study received ethical approval from a local Ethical Committee at the Faculty of Psychology and Neuroscience, Maastricht University. Completion of the program required that children attended at least 12 out of the 15 sessions. Nine children were excluded from the final data analyses due to parental refusal to sign the permission form or not attending enough sessions. For the ANT-I, the data of further 13 children was not usable due to technical problems with the equipment at pre-test and/or post-test sessions. After dropout, the data of 24 children was used in school 1 ( $N_{\text{mindfulness training group}} = 12$ ,  $N_{\text{active control group}} = 12$ ) and the data of 63 children was used from school 2 ( $N_{\text{mindfulness training group}} = 24$ ,  $N_{\text{active control group}} = 22$ ,  $N_{\text{passive control group}} = 21$ ).

### Design

The study used a non-randomized repeated-measures between-subjects design with one experimental group and one active control group in school 1 and one experimental, one active control group and one passive control group in school 2. Children were not randomly allocated to a group due to organizational constraints, so groups were determined by existing classes. One week before the start of training, participants and parents filled out the questionnaires for the pre-test, and within 2-3 weeks after completion of the training, post-test measurements were conducted.

## Measures and Materials

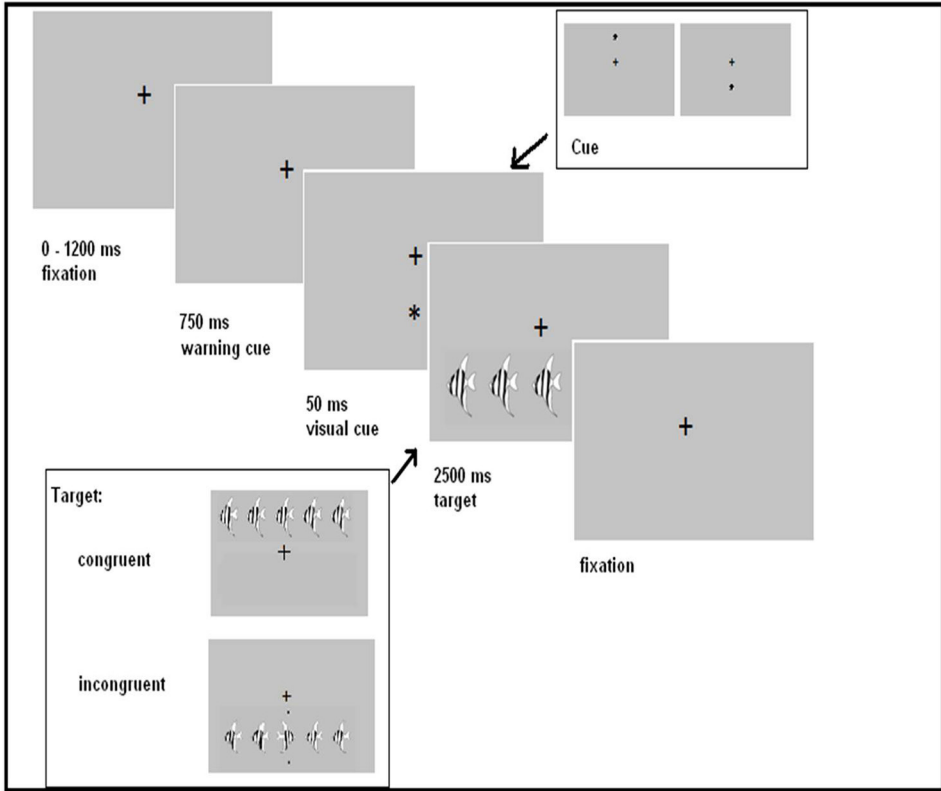
### *Attention Network Test (ANT-I)*

A previously adapted version of the Attention Network Test (ANT-I) (Callejas et al., 2004, 2005) was used to measure the capacities of the children's three attention networks (see Figure 1). This ANT-I differed in two regards to the ANT-I by Callejas and colleagues (2004, 2005): first of all, it used a more child friendly interface by using pictures of fishes instead of arrows. A similar design was used by Rueda and colleagues (2004) in their child-version of the ANT, and children were found to respond well to this design. Secondly, the stimulus onset asynchrony (SOA) between the alerting cue and the orienting cue was changed from 500 ms to 800 ms to account for children's slower processing times. The responses were collected via two keyboard keys (M and C). In every trial a fixation cross was held constant in the background. 24 trials were given in an initial practice round, followed by three blocks of 96 trials. After each block the child could briefly pause, and some form of encouragement ("well done" or "you're nearly there") was given.

The capacity of each network can be measured by calculating the mean Reaction Times (RT) in each condition and subtracting the relevant conditions from each other. Thus, the efficiency of the executive control network can be calculated by subtracting the mean RT across all congruent trials from the mean RT computed across all incongruent trials. The efficiency of the orienting network was inferred by calculating the difference between the mean RT across all validly cued target trials and the mean RT across all invalidly cued target trials (congruent and incongruent stimuli averaged). Finally, the effectiveness of the alerting network was determined by subtracting the mean RT on all uncued target trials from the mean RT on all cued target trials.

### *Attention Control Task*

The Attentional Control Task was conducted immediately after the ANT-I, when children had already experienced some depletion on their attention networks. In the eight-minutes attention control task, children were required to calculate and write down the sum of digits presented on the computer screen before them. The sums increased in difficulty throughout the task. Children were subjected, through headphones, to auditory interferences in the form of a woman's voice randomly reading two or three digit numbers between eleven and 170. Children could solve a total amount of 40 sums, which was also the maximum number of correct sums that children could obtain in this task. Those children who had not completed the ANT-I due to technical failures were not included in the analysis. This task was previously used by Alberts et al. (2007), although it was used as a depletion task and not as a dependent measure. While the task does not measure the efficiency of individual attention networks, it was included since it was found to have high ecological validity as children are subjected to a lot of auditory interference in class while having to concentrate on their school work. Furthermore, using the percentage of sums solved correctly while correcting for the amount of sums solved in total, a dependent measure was easily found.



**Figure 1:** Schematic of the child version of the amended version of the ANT (Fan et al., 2002). In the actual test, the background colour is white.

### *Behavior Rating Inventory of Executive Function (BRIEF)*

For all children, parental ratings of the Dutch BRIEF (Huizinga & Smidts, 2011) were taken at pre-test and at post-test. The BRIEF is a rating scale specifically developed to assess everyday executive skills of children aged five to eighteen in their daily environments such as home or school (Gioia et al., 2000). The parental version of the BRIEF has 75 questions using a three-point scale (Never, Sometimes, Often). 16 parents did not return the BRIEF at pre-training and the data was not included in the analysis. Since there were too few questionnaires returned by the parents at post-session it was not possible to analyze effects of the training on BRIEF-scores. The pre-training BRIEF scores on the Emotion Control were used to investigate whether emotion regulation behavior present before the training might influence training effects.

### *Procedure*

In School 1, training took place from February to April 2012 and in School 2 training took place from March to June 2012. In School 2 training was paused for two weeks

as the school closed for Easter break. Two to three weeks before the MT and reading group began, questionnaires were given to the parents for completion and were collected the week before training. During testing sessions, the children were individually taken to a secluded room. Once the ANT-I had been explained they were given a practice round. Only children who had six or less mistakes of a total of 24 trials were allowed to proceed with the ANT-I; the others had to practice the trials again until they reached a 72% correct criterion. Completion of the ANT-I took no longer than 25 minutes in total. After performance on the ANT-I, children were tested on the Attentional Control Task. At the post-test, the same procedure was repeated.

### *Treatment*

The experimental groups received the mindfulness teaching three times a week for 15 minutes for a five-week period. Thus, in total, children spent 45 minutes per week practicing mindfulness. Mindfulness training was given by trainers certified by Mindful Schools, an organization that adapted the MBSR course for children and that gives age-appropriate exercises and a fixed curriculum to their teachers.<sup>1</sup> A certified mindfulness teacher with at least two years experience in teaching Mindful Schools and four years experience in teaching Mindfulness taught the sessions to a whole class. During the course, children learned elements of FA meditation and OM meditation. Generally, the sessions would start with a Focused Attention meditation exercise (e.g. breathing with awareness) followed by theory about mindfulness, followed by an Open Meditation exercise. Afterwards, the children were asked to answer questions written on the board or reflect on the session by writing or drawing answers in a booklet. No homework was explicitly given, although children were encouraged to practice skills learned in the mindfulness group outside of sessions as well.

In the active control group, children were read a story from a book (School 1: "Artie in Artis" and "Dierenverhalen", School 2: "Artie in Artis" and "Otje"). Afterwards, the children reflected on the story or drew scenes from the book in their booklet. No homework was given in this group. No extra training was given for the passive control group in School B but the children participated in pre- and post-tests.

### **Statistical analysis**

The study comprised an active controlled repeated measures between subjects design in which different groups of children were assigned to either mindfulness training or active/passive control groups. Children's mean network scores (computed as reaction time difference scores) on the ANT-I, parent-rated scores on the BRIEF and amount correct sums on the Attention Control Task were the dependent variables. For the computation of network scores in the ANT-I, trials with missing or incorrect responses were excluded from the analysis, as were trials with a RT under 200 ms. This was due to the fact that 100 ms - 200 ms is the time needed to process and respond, so responses faster than this can be seen as the

---

1 A copy of the relevant course can be accessed on their webpage: [www.mindfulschools.org](http://www.mindfulschools.org).

result of fast guesses (Whelan, 2008). Since children have a slower processing time than adults, the upper limit (200 ms) was chosen. Efficiency of the executive attention, orienting and alerting networks before and after training was estimated by computing subsystem difference scores as described before (see 2.3.1). This is a common method of analysis and has been reported in other papers (see e.g. Jha et al., 2007 and Fan et al., 2002).

To evaluate effects of the Mindfulness Training on the functioning of the three attention networks, these ANT-I subsystem difference scores were entered in a univariate repeated measures analysis of covariance (ANCOVA). Separate analyses for executive control, orienting and alerting were conducted with a Group factor and a Time factor. The Group factor either consisted of two levels at school 1 (MT group and AC group) or of three levels at school 2 (including MT, AC and a third passive control (PC) group). Age was included as a covariate because of the possible influence of age differences on network maturity. We also investigated the possible role of differences in pre-training levels of emotion control on effects of the training on the ANT-I. To get enough power in this analysis, only children from active control groups and mindfulness training groups (collapsed across schools) were included. Children from both groups were divided in two groups of subjects with either low or high emotional control on the basis of a median split on BRIEF-Emotion Control scores obtained pre-training within each group. The influence of this between group emotion control factor on training effects on ANT-I scores was examined in an ANOVA with two between factors: EmotionControl (high/low) and TrainingGroup (MT and AC). Post – pre ANT-I difference scores for orienting, alerting and executive control were entered in the ANOVA.

To evaluate the effects of Mindfulness Training on the Attention Control Task, an ANCOVA with one between group factor (Amount of Correct Sums) and one within group factor TrainingGroup (MT and AC) was conducted. To get enough power for this analysis, only children from MT groups and active control groups (collapsed across schools) were included. To control for children's numerical skills, the amount of correct sums at pre-test was entered as a covariate.

## RESULTS

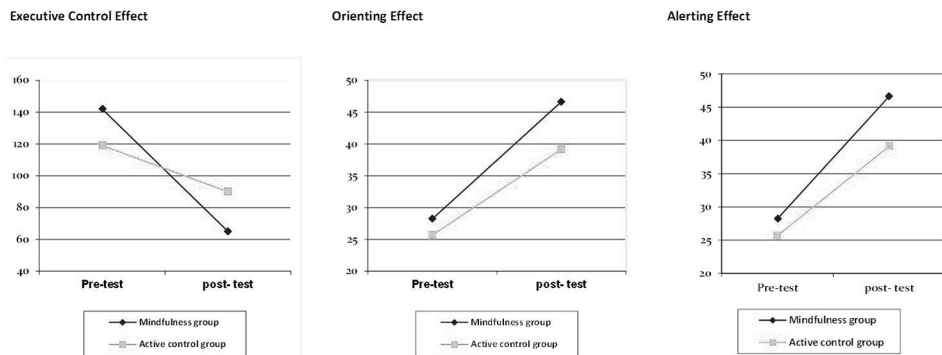
### **Effects of mindfulness training on Attention Network functioning**

#### *School 1:*

The means of the RTs for correct responses in each experimental condition and for the different groups and different schools are presented in Table 2 and Figure 2. Before conducting the analysis an independent samples t-test showed there were no significant differences between the groups at pre-test for Executive Effect ( $t(22) = .871, p = .393$ ), Orienting ( $t(22) = .221, p = .310$ ) and Alerting ( $t(22) = -1.328, p = .198$ ). It was thus decided to continue with the ANCOVA analysis. The ANCOVA analysis did not show a significant interaction of training and age. Further, there was no main effect of training group on mean RTs in Executive Control ( $F(1,21) = 2.605$ ,

$p=.121$ ), Orienting ( $F(1,21)=.139$ ,  $p=.713$ ), and Alerting ( $F(1,21)=.423$ ,  $p=.522$ ). Since age did not explain variance in training effects on subsystem functioning, an ANOVA was conducted without age. Although mean RT Executive Control scores improve more from pre- to post-training in the MT group than in the AC group (see Table 2 and Figure 2), no significant effect of Training Group on mean RT of Executive Control ( $F(1,22)=2.713$ ,  $p=.114$ ), Orienting ( $F(1,22)=.390$ ,  $p=.539$ ) or Alerting ( $F(1,22)=.880$ ,  $p=.358$ ) was found. There was a main effect of Time for Executive control ( $F(1,22)=13.350$ ,  $p=.0014$ ) indicating that Executive Control scores decreased from the pre to the post measurement. However, the lack of interaction with Group indicates that this decrease was present in trained and untrained groups and thus was not specific for the MT group. No main effects of Time for Orienting and Alerting were found.

To check whether the alerting, orienting and executive control manipulations of the ANT worked in the sense that there were significant task effects (independent of the training) it was checked whether there were main task effects (e.g. invalid-valid, no tone-tone, congruent-incongruent differences). Significant main task effects were present for Executive Control ( $F(1,22)=87.585$ ,  $p<.0001$ ), Orienting ( $F(1,22)=32.827$ ,  $p<.0001$ ) and Alerting ( $F(1,22)=79.363$ ,  $p<.0001$ ). P-values were rounded up to maximal 4 decimals behind the comma.



**Figure 2:** School 1 attention network effects in mean RT pre-test and post-test.

### School 2:

Figure 3 displays the mean changes in RT from pre-test to post-test in the three attention networks. Since there were no statistically significant differences between group means at pre-test as determined by one-way ANOVA for Executive Control ( $F(2,64) = 0.586$ ,  $p = .559$ ), Orienting ( $F(2,64) = 0.535$ ,  $p = .588$ ) or Alerting ( $F(2,64) = 0.496$ ,  $p = .611$ ), it was decided to continue with the ANCOVA analysis. No interaction with age was found in the ANCOVA analysis, and there was no main effect of training group on mean RTs in Executive Control ( $F(2,61)=1.822$ ,  $p=.170$ ), Orienting ( $F(2,62)=.092$ ,  $p=.913$ ), and Alerting ( $F(2,61)=.706$ ,  $p=.497$ ). As age did not explain variance in training effects on subsystem functioning an ANOVA was



conducted without age as a covariate. No significant effect of Training Group on mean RT of Executive Control ( $F(2,62)=1.880$ ,  $p=.161$ ), Orienting ( $F(2,63)=.090$ ,  $p=.914$ ) or Alerting ( $F(2,62)=.636$ ,  $p=.533$ ) was found. There was a main effect of Time for Executive control ( $F(1,62)=49.640$ ,  $p=.0001$ ) indicating that Executive Control scores decreased from the pre- to the post-measurement but lack of interaction with Group shows that it was not specific for the mindfulness group. No main effects of Time for Orienting and Alerting were found.

To check whether the alerting, orienting and executive control manipulations of the ANT worked, we analyzed whether there were main task effects. Significant main task effects were present for Executive Control ( $F(2,62)=349.720$ ,  $p<.0001$ ), Orienting ( $F(1,63)=22.659$ ,  $p<.0001$ ) and Alerting ( $F(1,62)=162.043$ ,  $p<.0001$ ). P-values were rounded up to maximal 4 decimals behind the comma.

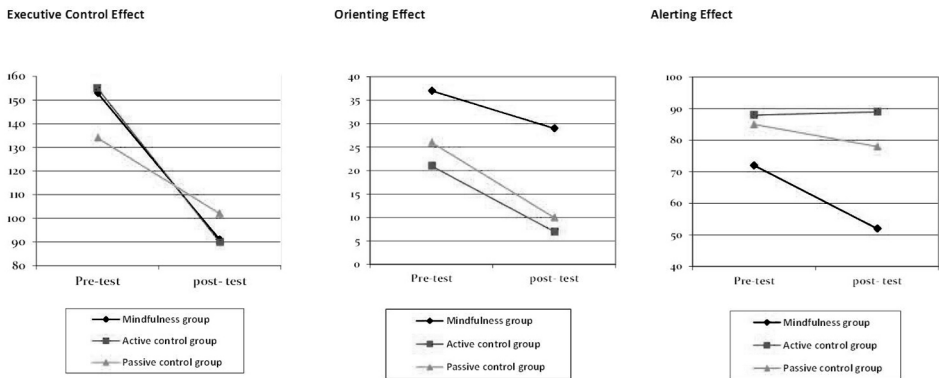


Figure 3: School 2 attention network effects in mean RT pre-test and post-test.

### Effects of Mindfulness Training on the Attention Control Task

The mean number of correctly solved sums before and after the training in MT groups and AC groups (averaged over the schools) are presented in Figure 4. An ANCOVA analysis with pre-test scores entered as covariate and post-test scores as dependent measure showed no main effect of TrainingGroup on Amount of Correct Sums at post-test ( $F(1,65)=1.276$ ,  $p=.263$ ).

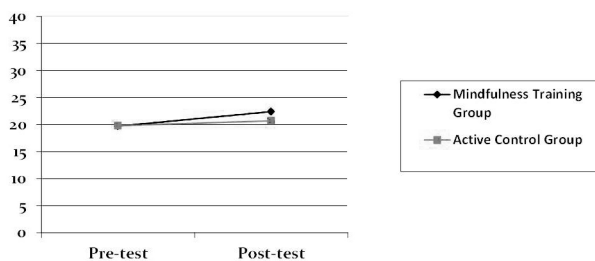


Figure 4: Mean amount of Correct Sums on the Attention Control Task from pre-test to post-test.



**Effects of pre-training differences in emotion control on training effects.**

A repeated measures ANOVA showed that no effects of pre-training Emotional Control levels (high/low) on effects of Training for the three subsystem attention network scores were present for Executive Control ( $F(1,54)=.138, p= .712$ ); Orienting ( $F(1,55)=1.273, p= .264$ ) and Alerting ( $F(1,54)=.197, p= .659$ ). No main effects were found for Emotion Control on the subsystem attention scores of the ANT: Executive Control ( $F(1,54)= .085, p= .772$ ), Orienting ( $F(1,55)= .240, p= .626$ ) and Alerting ( $F(1,54)= 1.252, p= .268$ ).

**DISCUSSION****Discussion of results**

The goal of this study was to investigate the effects of mindfulness training on children's attentional control. We hypothesized that the mindfulness manipulation group would improve their attentional performance compared to a reading group and a passive control group which was tested too but did not receive training. Furthermore, we hypothesized that this change would be particularly evident in the executive control and orienting network, with the mindfulness group's mean reaction times decreasing more than the mean reaction times in the control groups after training. Finally, we hypothesized that children with low emotional control would benefit more from mindfulness training than children with high emotional control.

Unfortunately, these hypotheses could not be validated as children in the mindfulness groups did not perform significantly better than children in the reading groups or passive control groups on the attention tasks. Additionally, children's emotional control at pre-test in the mindfulness group was not predictive of their attentional performance on the ANT-I after mindfulness training. This is a somewhat surprising result given the number of studies that report positive changes in both children and adult's attentional capacities after mindfulness training (for example Jha et al., 2007; Tang et al., 2007; Chambers et al., 2008; Baijal et al., 2011; Semple et al., 2010; Napoli et al., 2005). Several explanations for this, ranging from methodological differences in mindfulness trainings to psychometric challenges, will be discussed below.

First, it is important to note that while none of the training effects were statistically significant, especially for the attention measures with a higher appeal on Executive control, they do indicate a trend in the predicted direction for changes in executive control. In the first school, the Executive control network scores (e.g. reaction times) decrease more from pre- to post-training measurement in the mindfulness group than in the control group, whereas this is not the case for the Orienting and Alerting networks. A similar trend was found in the second school where mean Executive control RT scores decrease more after training in the mindfulness group compared to the passive control group (but here there were no differences with the active control group). Furthermore, with respect to the attentional control calculation task, although not significant, the children in the

MT group did show a slight increase in correct sums after the training, whereas the control group did not. Since previous studies also report largest mindfulness training effects in the executive control network, this seems indicative that our results are going in the right direction.

However, for the executive control network all groups show an improvement with repeated measurement, leaving the possibility that a large part of the effect is due to practice on the task: repeated exposure to the task may have simply allowed the children to get better at it. Yet this would not explain why a similar trend cannot be found in the alerting or the orienting networks. However, the lack of finding a significant interaction cannot be due to a lack of task effects in the ANT-I since significant task effects during pre- and post-measurement were found for all three networks.

There are several possible explanations for not finding evidence that mindfulness training supports children's attentional skills. First of all, compared to the "traditional" mindfulness training courses, the curriculum offered by Mindful Schools is very short. In MBSR trainings, for example, participants receive up to 72 hours of mindfulness while children participating in the Mindful Schools curriculum only receive a total of 3.75 hours of mindfulness sessions. Other studies that also employed a shorter exposure time of mindfulness have also reported little or no results. For example, Anderson et al. (2007) reported no relationships between mindfulness and measures of sustained attention, switching, elaborative processing, or non-directed attention in adults. The authors themselves suggested that a longer mindfulness training course may have been more effective in producing improvements of attentional control. Furthermore, given that some studies have found a correlation between higher network scores on the ANT-I and experience in meditation (Jha et al., 2007), the fact that not enough time was spent practicing mindfulness seems quite likely.

### **Psychometric Problems**

In the present study, the ANT-I was chosen as the instrument to measure attention since it is a commonly used instrument to distinguish three functionally different attention networks. Furthermore, prior studies have demonstrated its sensitivity for effects of brief behavioral interventions, such as Mindfulness trainings in adult populations (Tang et al. 2007). Rueda and colleagues reported in an article that previous studies had found that children respond well when there was a story and instant feedback, and had thus adapted their child-version of the ANT-I in this way (Berger, Jones, Rothbart, & Posner, 2000; Rueda et al., 2004). While the version of the ANT-I used in this study was presented to the children within a story (catching fishes), no feedback was given to the children beyond the practice round. This may have served as an additional motivator for children to answer questions more quickly, but it is likely that this would have been consistent in all groups, not just the mindfulness training group. Furthermore, there were highly significant task effects on reaction time for all three networks, showing that the ANT-I as a test worked well in our children. Finally, while this objection may explain some variation in the data it does not explain why the Attentional Control Task did not find significant results, and is thus not sufficient in itself to explain the above results.

### Other problems

Finally, questions have been raised about the relative age of the children. Baijal et al. (2011) comment that monastic traditions that use elements of Mindfulness Trainings do not offer formal meditation experience to young monks until they reach early adolescence. It is possible that the Indo-Tibetan monastic tradition has experienced fewer benefits in teaching younger children to meditate, allowing the teaching of meditation to coincide with the increased maturation of the frontal cortex. A pilot study by Semple, Reid and Miller (2005) indicated that mindfulness-based techniques could be taught to children as young as 7 years old, citing decreased anxiety scores as rated by the children's teachers and the school psychologist after training. However, since the study was done with only five children and looks at observer-rated emotion scores rather than neuropsychological tests it is not enough evidence to conclude that mindfulness is truly effective at a young age. The question of age is one that remains to be answered by future studies.

### CONCLUSION

Regardless of the above issues, this research raises a valid question whether mindfulness, in particular the Mindful Schools curriculum, is an effective way to support pre-adolescent children's attentional control. Given the amount of studies with adult populations reporting positive results, as well as a few children's studies which are certainly indicative of positive effects, it seems unlikely that children are completely unsusceptible to mindfulness training effects. One likely explanation for the results is the relative short time the children spent practicing mindfulness, compared to other studies which used longer training times and found positive changes in attentional control. Further research using a different mindfulness curriculum is recommended to fully assess the validity of this training.

### REFERENCES

- Alberts, H.J.E.M., Martijn, C., Greb, J., Merckelbach, H., & de Vries, N.K. (2007). Carrying on or giving in: the role of automatic processes in overcoming ego depletion. *British Journal of Social Psychology*, 46, 383-399.
- Anderson, N. D., Lau, M. A., Segal, Z. V., & Bishop, S. R. (2007). Mindfulness-based stress reduction and attentional control. *Clinical Psychology & Psychotherapy*, 14(6), 449-463.
- Arch, J. J., & Craske, M. G. (2006). Mechanisms of mindfulness: Emotion regulation following a focused breathing induction. *Behaviour Research and Therapy*, 44(12), 1849-1858.
- Baer, R. A. (2011). Measuring mindfulness. *Contemporary Buddhism*, 12(1), 241-261.
- Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report - The Kentucky inventory of mindfulness skills. *Assessment*, 11(3), 191-206.
- Baijal, S., Jha, A. P., Kiyonaga, A., Singh, R., & Srinivasan, N. (2011). The influence of concentrative meditation training on the development of attention networks during early adolescence. *Frontiers in Psychology*, 2, 153-163.
- Beauchemin, J., Hutchins, T. L., & Patterson, F. (2008). Mindfulness meditation may lessen

- anxiety, promote social skills, and improve academic performance among adolescents with learning disabilities. *Complementary Health Practice Review*, 13, 34-45.
- Berger, A., Jones, L., Rothbart, M.K., & Posner, M.I. (2000). Computerized games to study the development of attention in childhood. *Behav Res Methods Instrum Comput*, 32(2), 297-303.
- Biegel, G. M., Brown, K. W., Shapiro, S. L., & Schubert, C. A. (2009). Mindfulness-Based Stress Reduction for the treatment of adolescent psychiatric outpatients: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 77(5), 855-866.
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, 78(2), 647-663.
- Boegels, S., Hoogstad, B., van Dun, L., de Schutter, S., & Restifo, K. (2008). Mindfulness training for adolescents with externalizing disorders and their parents. *Behavioural and Cognitive Psychotherapy*, 36(2), 193-209.
- Callejas, A., Lupiáñez, J., Funes, M. J. s., & Tudela, P. (2004). The three attentional networks: on their independence and interactions. *Brain and Cognition*, 54, 225-227.
- Callejas, A., Lupiáñez, J., Funes, M. J. s., & Tudela, P. (2005). Modulations among the alerting, orienting and executive control networks.
- Chambers, R., & Allen, N. B. (2005). The effects of mindfulness meditation on psychological wellbeing and executive control of attention. *Australian Journal of Psychology*, 57, 21-21.
- Chambers, R., Lo, B. C. Y., & Allen, N. B. (2008). The impact of intensive mindfulness training on attentional control, cognitive style, and affect. *Cognitive Therapy and Research*, 32(3), 303-322.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social-information processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115, 74-101.
- Ehrenreich, J. T., & Gross, A. M. (2002). Biased attentional behavior in childhood anxiety: A review of theory and current empirical investigation. *Clinical Psychology Review*, 22(7), 991-1008.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Cognitive Neuroscience*, 14(3), 340-347.
- Flook, L., Smalley, S. L., Kitil, M. J., Galla, B. M., Kaiser-Greenland, S., Locke, J., et al. (2010). Effects of mindful awareness practices on executive functions in elementary school children. *Journal of Applied School Psychology*, 26, 70-95.
- Gioia, G.A., Isquith, P.K., Guy, S.C., & Kenworthy, L. (2000). *Behaviour Rating of Executive Function*. Lutz, FL: Psychological Assessment Resources.
- Huang-Pollock, C., Mikamo, A. Y., Piffner, L., & McBurnett, K. (2007). ADHD Subtype differences in motivational responsivity but not inhibitory control: Evidence from a reward-based variation of the stop signal paradigm. *Journal of Clinical Child and Adolescent Psychology*, 36(2), 127-136.
- Hughes, C., Cutting, A., & Dunn, J. (2001). Acting nasty in the face of failure? Longitudinal observations of "Hard-to-Manage" children playing a rigged competitive game with a friend. *Journal of Abnormal Child Psychology*, 29(5), 403-416.
- Huizinga, M. & Smidts, D.P. (2011). Age-related changes in executive function: a normative study with the Dutch version of the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology*, 17(1), 51-66.
- Jetten, G.E.J. (2012). *Attention Development & Autism Spectrum Disorder: Do the three attention networks interact? A behavioural study*. (Unpublished master thesis). Maastricht University, Maastricht.
- Jha, A. P., Krompinger, J., & Baime, M. J. (2007). Mindfulness training modifies subsystems of attention. *Cognitive Affective & Behavioral Neuroscience*, 7(2), 109-119.
- Kabat-Zinn, J. (1994). *Wherever you go, there you are: mindfulness meditation in everyday life*. New York: Hyperion.
- Kerns, K.A., Esso, K. & Thompson, J. (1999). Investigation of a direct intervention for improving

- attention in young children with ADHD. *Developmental Neuropsychology*, 16, 273–295.
- Kindt, M., Brosschot, J. F., & Everaerd, W. (1997). Cognitive processing bias of children in a real life stress situation and a neutral situation. *Journal of Experimental Child Psychology* 64, 79–97.
- Martin, M., Horder, P. & Jones, G. V. (1992). Integral bias of naming in phobia-related words. *Cognition and Emotion*, 6(6), 479–486.
- Napoli, M., Krech, P. R., & Holley, L.C. (2005). Mindfulness training for elementary school students: The attention academy. *Journal of Applied School Psychology*, 21, 99–125.
- Pashler, H., Johnston, J.C. & Ruthruff, E. (2001). Attention and performance. *Annual Review Psychology*, 52, 629–51.
- Posner, M.I. (1978). *Chronometric Explorations of Mind*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Posner, M. I. (2008). Measuring alertness. *Ann. N. Y. Acad. Sci.*, 1129, 193–199.
- Posner, M. I., & Boies, S.J. (1971). Components of attention. *Psychological Review*, 78(5), 391–408.
- Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Riggs, N., Blair, C., & Greenberg, M. (2003). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology*, 9, 267–276.
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J., Gruber, D. B., Lercari, L. P., et al. (2004). Development of attentional networks in childhood. *Psychologia*, 42(8), 1029–1040.
- Semple, R. J., Lee, J., Rosa, D., & Miller, L. F. (2010). A randomized trial of Mindfulness-Based Cognitive Therapy for children: Promoting mindful attention to enhance social-emotional resiliency in Children. *Journal of Child and Family Studies*, 19(2), 218–229.
- Semple, R. J., Reid, E. F., & Miller, L. (2005). Treating anxiety with mindfulness: An open trial of mindfulness trianing for anxious children. . *Journal of Cognitive Psychotherapy*, 19(4), 379–392.
- Tang, Y.Y., Ma, Y., Wang, J., Yaxin, F., Shigang, F., Lu, Q., Yu, Q., Sui, D., Rothbart, M.K., Fan, M. & Posner, M.I. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), 17152–17156.
- Taghavi, M., Neshat-Doost, H., Moradi, A. R., Yule, W., & Dalgleish, T. (1999). Biases in visual attention in children and adolescents with clinical anxiety and mixed anxiety-depression. *Journal of Abnormal Psychology*, 27(3), 215–223.
- Thorell, L. B., Lindqvist, S., Bergman, N., Bohlin, G., & Klingberg, T. (2008). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12(1), 106–113.
- Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, 73(3), 282–299.
- Trick, L. M., & Enns, J. T. (1998). Lifespan changes in attention: The visual search task. *Cognitive Development*, 13(3), 369–386.
- Vasey, M. W., Daleiden, E. L., Williams, L. L., & Brown, L. (1995). Biased attention in childhood anxiety disorders: a preliminary study. *Journal of Abnormal Child Psychology*, 23, 267–279.
- Vasey, M. W., El-Hag, N., & Daleiden, E. L. (1996). Anxiety and the processing of emotionally threatening stimuli: distinctive patterns of selective attention among high- and low-test-anxious children. *Child Development*, 67, 1173–1185.
- Whelan, R. (2008). Efective analysis of reaction time data. *The Psychological Record*, 58, 475–482.
- Zylowska L., Ackerman D.L., Yang M.H., Futrell, J.L., Horton N.I., Hale T.S., Hale, T.S., Pataki, C. & Smalley, S.L. (2008). Mindfulness meditation training in adults and adolescents with ADHD: A feasibility study. *Journal of Attention Disorder*, 11(6), 737–746.

# Driving home for Christmas: Influences of music tempo and inhibition training on simulated driving performance

## ORIGINAL PAPER

In modern society, a car is among the most used means of transportation. The amount of car accidents that involve young drivers increases every year and poses a serious societal problem in terms of personal, social and economic costs. An explanation for these accidents is given by a biological theory, which states that an immature prefrontal cortex results in riskier behaviour. The socio-environmental theory indicates environmental factors, such as peer pressure and education, as possible determinants of the increased risk in young drivers. The current study combines both theories by searching an effect of music tempo (environment) and response inhibition (biological) on driving performance. The results showed a main effect for impulsivity/inhibition on crash rates. This suggested that impulsive behaviour promoted focused attention, thus leading to a lower crash rate. An interaction effect showed a marginal music tempo effect on lap times, but only when showing impulsive behaviour.

**Keywords:** music tempo; response inhibition; driving

Chris Makkinje; Master Student Neuropsychology  
University Maastricht, Maastricht, the Netherlands

---

[ccc.makkinje@student.maastrichtuniversity.nl](mailto:ccc.makkinje@student.maastrichtuniversity.nl)

“Risk comes from not knowing what you’re doing.” —Warren Buffett  
(as cited in *Warren Buffet says*, 2009, p. 1).



## INTRODUCTION

Risky behaviour in adolescence and young adulthood is a well-known and studied subject (Arnett, 1992; Casey, Getz, & Galvan, 2008; Jessor, 1992). Through years of research, different theories have risen to explain the nature of this behaviour and why it decreases on the road to adulthood. The explanations of these theories can be sorted in two main streams: The biological and socio-environmental explanations.

A *biological* explanation was drawn from the fact that during adolescence the prefrontal and striatal systems go through developmental changes which are linked to risky behaviour (Casey et al., 2008; Casey et al., 2010; Galvan, Hare, Voss, Glover, & Casey, 2007; Steinberg, 2008b). In accordance to these developmental changes, a dual process theory was formed. This theory states that the pre-frontal cortex is responsible for controlled and rational behaviour and the limbic system is responsible for intuitive and uncontrollable behaviour (Rivers, Reyna, & Mills, 2008; Steinberg, 2008a). A new theory was derived from the dual process theory in which an explanation was given for the risky behaviour shown by adolescents and young adults. Casey et al. (2010) called it the imbalance model. In this model, a development trajectory was proposed in which the maturation of the subcortical limbic system develops faster than that of the pre-frontal cortex. According to this model, risky behaviour is more likely to occur in adolescence because of the inability of the pre-frontal cortex to sufficiently control impulsive responses to rewarding stimuli due to an oversensitive limbic system.

A *socio-environmental* explanation for increased risk behaviour in adolescents was given by the Broad and Narrow Socialization Theory of Arnett (1995). This theory emphasizes the importance of the changes in the sociocultural environment and states that risky behaviour is derived from these changes. Peer influence for example is a supporting factor for this theory. It has been proven by various studies that peer influence can have great effects on risky behaviour. Peers have an influence on this type of behaviour through direct peer pressure or by indirect pressure via socialization effects (Gardner & Steinberg, 2005; Maxwell, 2002). One of these effects is the tendency of adolescents to modify their behaviour and attitude so it will resemble those of friends (Brechwald & Prinstein, 2011).

Driving is a scenario in which risk-taking behaviour of young adults poses a serious threat to society. In a study by Arnett, Offer, and Fine (1997), 139 high school students received a questionnaire about reckless driving. About 80% of the males and 70% of the females reported to engage in this kind of behaviour in the form of car racing, driving over a no-passing zone and driving intoxicated. This supports the conclusion of risky driving behaviour by adolescents and young adults.

A socio-environmental factor that might influence risky driving behaviour is the use of music. According to Dibben and Williamson (2007), over 75% of the British population engages in some sort of music listening while driving. To understand the effect of music while driving, music itself can be divided in different factors that might have an influence on driving behaviour. Music volume, or in other words, intensity is one of those factors. North and Hargreaves (1997) revealed that drivers who are caught up in heavy traffic will be more inclined to decrease the music volume. Music will be decreased because the loudness itself requires more

processing (North & Hargreaves, 1997). This would lead to a cognitive overload and would result in a higher accident rate. Beh and Hirst (1999) investigated the effects of loud music on driving performance. They wanted to examine whether loudness had a different effect on low demanding or high demanding driving tasks. In this experiment participants had to drive with either soft music (55 dBA) or loud music (85 dBA). The driving was done under two conditions: low-demanding single-task driving and high-demanding multi-task driving. The authors found that soft music did not show an effect on the reaction time during driving. However, louder music interfered significantly with the reaction time especially in the high-demanding task.

Besides music intensity, music complexity also has an influence on driving behaviour. North and Hargreaves (1999) conducted an experiment on the effects of music complexity while driving. They hypothesized that high arousing music would lead to slower lap times in comparison to low arousing music due to an overload of cognitive functioning and therefore impaired driving skills. In their study, 5 laps had to be raced in a computerized environment. These laps were conducted under influence of either a low demanding driver task or a high demanding driver task. Combined with the driver task, either low arousal music (80 BPM at 60 dBA) or high arousal music (140 BPM at 80 dBA) was presented. In their results it was concluded that the worst lap times coincided with presentation of high arousing music and a high demanding driver task, thus providing support for the cognitive overload hypothesis.

In the previous studies both music intensity and complexity could influence driving behaviour. A third possible factor that is entwined within these studies is musical tempo. Fast tempo music has shown to lead to higher arousal levels (van der Zwaag et al., 2011). For example, the presentation of fast-paced music led to significant faster shopping in the supermarkets and caused faster drinking and eating (Herrington, 1996; McElrea & Standing, 1992). An effect of musical tempo in driving has been shown by Brodsky (2001). He discussed that arousing music was more cognitively demanding than non-arousing music because of a larger number of temporal events (for example more beats in a shorter period of time). Therefore higher music tempo could lead to a cognitive overload. In the experiment, it was hypothesized that fast paced music would have an effect on simulated race driving reflected in decreasing simulated lap-time (measured in accelerated speeds) and an increase traffic violations. These outcomes would reflect faster driving and more reckless behaviour. In this study subjects had to drive eight laps of six miles each through the city of Chicago. For this experiment the computer game 'Mid town madness' was used, which provided not only traffic but also traffic rules (stop signs, traffic lights etc.). Each of the eight laps was divided into three zones, of which two were accompanied by slow, medium or high tempo music. In his results, Brodsky found a main effect for tempo. Participants drove faster and caused more traffic violations when presented with fast-tempo music.

In conclusion, music has shown in many ways that it can have an influence on driving behaviour. In the studies of Brodsky (2002) and North and Hargreaves (1999), this influence was hypothesized to result from a high arousal effect of the music. This arousal is seen as a predictor of impulsive behaviour and therefore risky



behaviour (Anderson & Revelle, 1994; Mattila & Wirtz, 2001).

A biological factor that might influence risky driving behaviour is the development of response inhibition. According to the biological explanation, risky behaviour lies within an immature pre-frontal cortex. One of the functions of the pre-frontal cortex, which has been researched for years, is executive functioning. A lot of this research has been reviewed by Pennington and Ozonoff (1996). According to their review, executive functioning selects appropriate actions in each situation, depending on the context. The most important functions that lie within the executive functions are: set-shifting, planning, working memory, and response inhibition. The latter is one of the most important functions since response inhibition facilitates the suppression of actions that are not of any use and actions that do not fit in the context. Therefore it is essential for goal-directed behaviour, and needs to be adaptable to many different contexts. Driving performance is a highly goal-directed task, which requires adequate functioning of the prefrontal cortex. An immature prefrontal cortex and consequently not properly developed response inhibition may produce impairing effects on driving performance (Jongen, Brijs, Komlos, Brijs, & Wets, 2011). In a study by Jongen et al. (2011) this inhibitory control was measured via a stop signal reaction task. In this task, a participant had to respond to a stimulus and inhibit this response whenever a sound was given prior to this stimulus. It was hypothesized that inhibitory control would improve with age and therefore improve driving skills. Subjects were divided into two groups based on age. The first group consisted of participants aged 17 to 18 and a second group aged 22 to 24. Both groups had no more than 2 years of driving experience. They concluded that the first group had lower inhibitory control, due to an immature pre-frontal cortex. This resulted in riskier driving skills, measured via different methods such as speeding, red light running and number of collisions.

The present study tries to combine both the biological- and the socio-environmental perspective. It has been discussed how music could influence driving via three main factors: music intensity, complexity and tempo. The first two factors have been investigated in quite a large extent compared to the latter. Therefore music tempo is used as a socio-environmental factor to investigate its effect on driving. In this study a racing game is used to simulate driving, and the lap time and the number of crashes is being used as variables to measure risky driving. The hypothesis is that the lap times will become faster and crashes will be more frequent as the tempo of the music is raised.

It has also been discussed how immature response inhibition can have a great influence on driving behaviour. In this study a stop signal task is used in order to elicit an inhibitory state or an impulsive state. According to a study by Guerrieri, Nederkoorn, Schrooten, Martijn, and Jansen (2009), using this task before the testing procedure and not as a post-condition measurement tool, would lead to an inhibitive state. They reasoned that the response inhibition learned in this task would reduce impulsivity. Furthermore, Guerrieri et al. (2009) pose that the original stop signal task can be altered to elicit an impulsive state. This state can be achieved by ignoring the no-response stimuli and focussing mainly on quick responding. The current hypothesis is that lap times will be slower and crashes will be less frequent after having conducted the response inhibition task as opposed to the impulsive

task. An interaction is predicted between impulsivity/inhibition and music tempo. This third hypothesis states that response inhibition will not lead to an effect of music tempo in contrast to response impulsivity, which might even facilitate the effect of music tempo.

## METHODS

### Participants

This study was approved by the psychology ethical committee at Maastricht University. Participants were deemed eligible if they were between 18 and 22 years old at the time of testing. This study consisted of 9 male and 25 female subjects. Their age varied from 18 to 22 years, ( $M = 20.23$ ;  $SD = 1.07$ ). All the subjects were psychology undergraduates and they received credit points in return for their participation.

### Design

This study used a  $3 \times 2$  design in which all participants were randomly assigned. The experiment varied the tempo of the music (Slow tempo, Normal tempo, High tempo) as a within subject factor and inhibitory control (inhibition vs. impulsivity) as a between-subject factor.

### Materials

The music that was used in this research had to be played with a few factors in mind such as the complexity and pitch changes. The song Nara of E.S Posthumus was selected because it did not have a high complexity nor did it have high pitch changeability. These factors could both act as a confounder. It was played in every condition with a headphone and connected to a Hp pavilion dv5 laptop. The beats per minute, 60 BPM, 100 BPM or 130 BPM, could be changed throughout the entire experiment using Virtual DJ Pro. The constant intensity of  $\pm 85$  decibel (Brodsky, 2002) was measured with a calibrated decibel meter. The stop signal task, which was used to evoke inhibition or impulsive behaviour, was presented via the program *presentation*® on a Dell Intel core 2 quad q8400 computer.

The simulator consisted of a Samsung ps-42c91h plasma TV connected to an Intel core 2 duo e8200 computer. The driving material used consisted of a bucket seat and Logitech gearbox, steering wheel, sound system and pedals. The game used for this experiment was *Racedriver Grid*. This game allowed the use of the simulator and it also gave the possibility to change the sensitivity of the pedals and the steering wheel. The simulator was used because the settings of these materials ensured the resemblance of driving in real life.

## Measures

### *Sensation Seeking Scale*

Zuckerman, Kolin, Price, and Zoob (1964) developed a questionnaire, the sensation seeking scale. This scale is used to predict sensation-seeking behaviour. The SSS5 consists of 40 questions posed in an A vs. B manner. These questions represent the 4 factors of this scale: thrill and adventure seeking, experience seeking, disinhibition and boredom susceptibility (Zuckerman, 1994).

### *Barratt Impulsivity Scale*

The BIS-11 scale consists of thirty questions rated on a 4-point scale (rarely/never, occasionally, often, and almost always/always). These questions reflect 3 different sub traits of impulsiveness: motor impulsiveness, additional impulsiveness and non-planning impulsiveness (respectively acting without thinking, not focusing on a task and a lack of future orientation) (Guerrieri et al., 2009; Patton, Stanford, & Barratt, 1995).

### *Stop signal task*

The stop signal task was created by Logan, Schachar, and Tannock (1997) as a way to measure impulsive behaviour in terms of a lack in response inhibition. This task consisted of 96 trials in each of the 4 blocks. The task consisted of 2 parts; first a choice reaction task (go-signals), in which the subjects were told to press the button related to the stimulus as fast as possible (in this case X or O). These stimuli were presented on a computer screen and disappeared when the right answer was given. After a few go signals, the same trial was given but with an external sound, cueing the participants to inhibit their response. In 25% of all the trials, this inhibition cue was given. Before this task, the participant's attention was focused on the importance of inhibiting the stimuli when cued by an external sound. According to Guerrieri et al. (2009), this focus should lead to an increase in response inhibition, and therefore inducing an inhibitive state.

Furthermore, Guerrieri et al. (2009) argued that it was possible to elicit impulsive behaviour by adapting the stop-signal task. In this adapted task, only the instructions were changed. The task itself remained the same as the one created by Logan (1997). The participants were told to ignore the inhibition cue and focus on the task as if the whole task consisted of only go signals. By treating the task as if it was a choice reaction task without inhibition cues, impulsive behaviour was elicited by focussing on the speed of this task.

Both tasks were used in this experiment in 2 separate conditions to obtain either an inhibitory or impulsive state.

## Procedure

Upon entering the room the subjects were asked to read and accept the informed consent. This was followed by a test with the headphones. Normal tempo music was played for a brief moment of approximately 10 seconds at the same volume

as it would be in the experiment. This was done to ensure that there wouldn't be any dropouts due to the loud music. The participants were then seated in front of the stop-signal task computer and were asked to fill in the BIS<sub>11</sub>, SSS<sub>5</sub> and some demographic questions.

When the questionnaires were completed, participants were placed inside the simulator for a practice trial in game mode with the same settings as there would be in the experiment. During the 5 laps of practice, the experimenter followed a memorized script in which several important things were explained to the subject such as how to drive through corners, the slippery effect of track side sand and grass, and how to steer properly.

As the practice session ended, the participant was instructed to take place behind the task computer. Participants were randomly assigned to the inhibition task or impulsive task. After a short practice task of 29 trials in which no inhibition cue was given, subjects performed either the inhibition or impulsive task for 15 minutes.

After the stop-signal task was performed, the subjects were asked to take seat in the driving simulator. The instructions for the driving experiment were given by the experimenter to execute the experiment properly. These instructions were always the same. After the three races, participants were asked to complete a last questionnaire on the stop-signal task computer. In this questionnaire musical preference and the purpose of this study was asked according to their own opinion.

A few weeks after the experiment was finished and all the data was collected, a debriefing was sent to all participants in which the true purpose was revealed.

### **Statistical analysis**

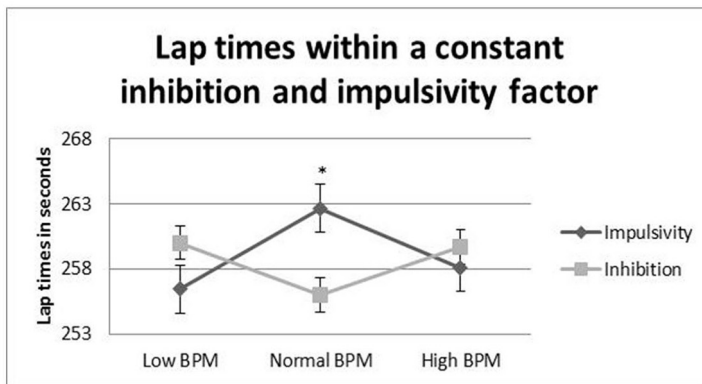
All obtained data were analysed using IBM SPSS 19<sup>\*</sup> with a repeated measures ANOVA on corrected lap times and crashes. The beats per minute were used as within subject factor and inhibition vs. impulsivity as between subject factors. Sex was not used as a between subject factor due to an imbalance of women to men ratio. The BIS and SSS scale were used as covariates and were centered due to a high correlation with the dependent variable. Age and driving experience were not further used during analysis as they proved to be non-significant when used as covariates.

## **RESULTS**

### **Lap times**

Lap times appeared not to be different between the three sessions with different music tempo,  $F(2, 29) = .174, p = .841$ . In addition, lap times did not differ between the group performing the inhibition task and the group performing the impulsivity task,  $F(1) = .686, p = .414$ . However, a significant Inhibition by Music Tempo interaction effect was found,  $F(2, 29) = 4.334, p = .023$ .

The interaction effect was further analysed with the use of simple effects. A marginal significant effect for Music Tempo was found in the impulsivity condition,  $F(2, 13) = 2.492$ ,  $p = .101$ . This effect was not found in the inhibition condition,  $F(2, 13) = 1.812$ ,  $p = .202$ , indicating that Music Tempo seems to have a stronger effect on impulsivity when participants were not asked to inhibit their responses on the computer task. As a post hoc testing, paired sampled t-tests were conducted on the lap times in the impulsivity condition. A significant effect was found between the low tempo music and the normal tempo music,  $t(16) = -2.387$ ,  $p = .03$ , and a trend was found between normal tempo music and high tempo music,  $t(16) = 1.475$ ,  $p = .16$ .

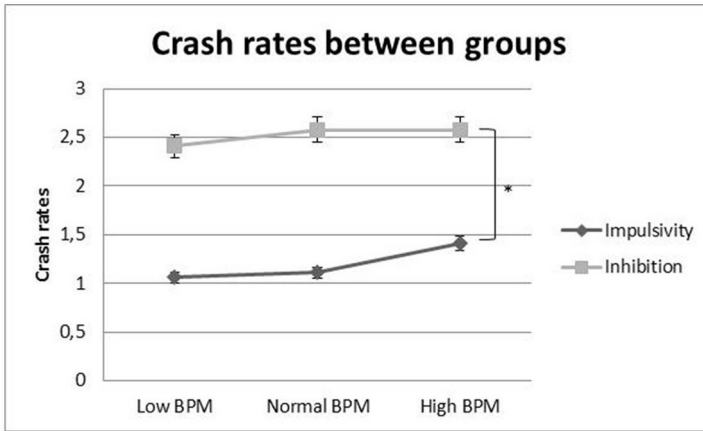


**Figure 1:** Lap times on the different music conditions while the inhibition and impulsivity factor were kept constant. A significant effect was found in the impulsivity condition between Low and Normal BPM. \* =  $p < 0.05$

### Crash rates

With the use of the same method, the crash rates between the three conditions were compared. There appeared to be no main effect for Music Tempo on crash rates,  $F(2, 29) = .316$ ,  $p = .732$ . In addition there was no interaction effect between BPM and inhibition,  $F(2, 29) = .265$ ,  $p = .769$ .

However a significant main effect for inhibition vs. impulsivity was found,  $F(1) = 6.953$ ,  $p = .013$ , showing that across music tempo conditions, those in the impulsivity condition had less crashes compared to those in the inhibition condition (see Figure 2).



**Figure 2:** The crash rates for inhibition and impulsivity. This figure shows a main effect between impulsivity and inhibition. \* =  $p < 0.05$

## DISCUSSION

The current study investigated possible factors that might influence driving behaviour. Driving is a high demanding task in everyday life, which can be influenced in a good or bad manner. In this study, music was used as a factor that could influence driving behaviour in a negative way, since it contributes to the cognitive load. Response inhibition versus impulsivity was used as a factor that might facilitate driving because a calm and inhibitive state might aid in driving. A driving simulator was used in combination with different music tempo and an adapted stop-signal reaction task to create an environment in which these factors could be measured.

In this study two main hypotheses and an interaction hypothesis were formed regarding the influence of these factors. Those could be measured via lap times and crash rates because these variables lend themselves in a race game. The first hypothesis investigated the effect of music tempo by posing that crash rates would increase and lap times would decrease in response to higher tempo of the music. The second hypothesis investigated the effect of inhibition versus impulsivity by posing that inhibition would lead to lower crash rates and slower lap times.

Studies conducted by Brodsky (2002) and North and Hargreaves (1999) revealed a musical tempo effect for driving speeds within a race game. The results of this study are not completely in line with those of previous studies. A main effect of music tempo was not found for lap times nor was it found for crash rates. The results show a non-significant, though increasing trend in crash rate in association with increasing music tempo. This appears to indicate that the results of the present study corroborate with previous findings. Probable power issues may explain the lack of significance.

The combined results of Logan et al. (1997) and Guerrieri et al. (2009) implied the possibility to elicit either an inhibitory mental state or an impulsive mental state by using a modified inhibition task. The results of this study did not show a main effect for inhibition vs. impulsivity on lap times but it did for crash rates. The hypothesis states that crash rates would be lower if inhibition was used. The results derived for crash rates showed a significant main effect in the opposite direction. Crash rates were lower in an impulsive mental state as opposed to an inhibitive mental state. This might be explained by posing that impulsivity would lead to a more focussed mental state that would aid in difficult situations, such as a more difficult corner in the game. Another explanation could be given by assuming that this effect would be the opposite when compared in real driving because of a generalization problem of race games with respect to driving in a real life situation. The generalization problem would occur because of the complexity of decisions that need to be made in a crowded environment, instead of a relatively simple estimation of when to steer in difficult corners.

An interesting result that this study found was an interaction effect between music tempo and inhibition vs. impulsivity. A marginal significant main effect of music tempo was found only in the impulsivity condition. The main effect was further investigated by a post hoc test, which revealed a significant effect for low tempo compared with normal tempo and a marginal significance for normal vs. high tempo music. However, these results are not entirely in line with the expectations because it seems that low tempo music reduces lap times. An explanation could be that when in an impulsive state, low tempo, and therefore low arousing music, would interfere less with driving resulting in better lap times. In conclusion, these results aid to the hypothesis of an effect of music tempo that was masked in this experiment due to the mental state factor. An explanation could be that inhibition training does have an effect on driving behaviour by inhibiting the effects of music tempo.

In order to interpret these results correctly, the limitations of this study need to be considered. One major limitation of the study was that there were bits of subjectivity that were impossible to erase. The lap times used in the analyses were corrected for crashes. So it could be possible to obtain slightly different lap times due to these time corrections. This might influence the effects of both inhibition and music. Another limitation was the use of a race game instead of the driving software. Although a driving simulator was used, a race game may not be entirely generalizable in every day driving due to the complexity of real driving scenarios. It could also be possible that there were small flaws in the design of the study. For example, the presence of the experimenter during the race could interfere because of an unintentional attention shift during driving. The use of headphones for presenting the auditory stimuli could also have an effect because it is unnatural to drive with headphones. The amount of participants could also have great effects on this study. If a higher number of participants could be tested, it might alter the results to find more significant effects and therefore might give more conclusive results in relation to the hypotheses that were made. A final major limitation was the use of the stop signal task. Guerrieri et al. (2009) showed that these tests could be used to induce either an inhibitive or an impulsive state. However, the duration



of these effects have not been studied as of yet. So it could be that these effects were not long lasting enough through the entire experiment.

In conclusion, the current study provides some insights into the role of music while driving and people's capacity to mitigate possible negative influences of music. Even though a conclusive result cannot be given, it seems that music effects are present during driving and especially high tempo music results in faster driving and a trend in higher crash rates which implies that high tempo music contributes to riskier driving.

This study also examined the relation between an impulsive and an inhibitive mental state. Even though no main effects were found, interaction between these factors and BPM suggests an inhibitive effect on music tempo.

### Acknowledgements

I would like to thank my two supervisors, Loes Kessels and Rob Ruiter, for their help during the research and for granting me their permission to publish this study.

### REFERENCES

- Anderson, K. J., & Revelle, W. (1994). Impulsivity and time of day: Is rate of change in arousal a function of impulsivity? *Journal of Personality and Social Psychology*, 67(2), 334-344.
- Arnett, J. J. (1992). Reckless behavior in adolescence: A developmental perspective. *Developmental Review*, 12(4), 339-373.
- Arnett, J. J. (1995). Broad and narrow socialization: The family in the context of a cultural theory. *Journal of Marriage and the Family*, 57(3), 617-628.
- Arnett, J. J., Offer, D., & Fine, M. A. (1997). Reckless driving in adolescence: "State" and "trait" factors. *Accident Analysis & Prevention*, 29(1), 57-63.
- Beh, H. C., & Hirst, R. (1999). Performance on driving-related tasks during music. *Ergonomics*, 42(8), 1087-1098.
- Brechwald, W. A., & Prinstein, M. J. (2011). Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes. *Journal of Research on Adolescence*, 21(1), 166-179. doi: 10.1111/j.1532-7795.2010.00721.x
- Brodsky, W. (2001). The effects of music tempo on simulated driving performance and vehicular control. *Transportation Research Part F: Traffic Psychology and Behaviour*, 4(4), 219-241.
- Casey, B. J., Getz, S., & Galvan, A. (2008). The adolescent brain. *Developmental Review*, 28(1), 62-77.
- Casey, B. J., Jones, R. M., Levita, L., Libby, V., Pattwell, S. S., Ruberry, E. J., . . . Somerville, L. H. (2010). The storm and stress of adolescence: Insights from human imaging and mouse genetics. *Developmental Psychobiology*, 52(3), 225-235.
- Dibben, N., & Williamson, V. J. (2007). An exploratory survey of in-vehicle music listening. *Psychology of Music*, 35(4), 571-589.
- Galvan, A., Hare, T., Voss, H., Glover, G., & Casey, B. J. (2007). Risk-taking and the adolescent brain: Who is at risk? *Developmental Science*, 10(2), F8-F14.
- Gardner, M., & Steinberg, L. (2005). Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: An Experimental Study. *Developmental Psychology*, 41(4), 625-635.
- Guerrieri, R., Nederkoorn, C., Schrooten, M., Martijn, C., & Jansen, A. (2009). Inducing impulsivity leads high and low restrained eaters into overeating, whereas current dieters



- stick to their diet. *Appetite*, 53(1), 93-100.
- Herrington, J. D. (1996). Effects of music in service environments: a field study. *Journal of Services Marketing*, 10(2), 26 - 41.
- Jessor, R. (1992). Risk behavior in adolescence: A psychosocial framework for understanding and action. *Developmental Review*, 12(4), 374-390. doi: [http://dx.doi.org/10.1016/0273-2297\(92\)90014-S](http://dx.doi.org/10.1016/0273-2297(92)90014-S)
- Jongen, E. M. M., Brijs, K., Komlos, M., Brijs, T., & Wets, G. (2011). Inhibitory control and reward predict risky driving in young novice drivers – a simulator study. *Procedia - Social and Behavioral Sciences*, 20(0), 604-612.
- Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science*, 8(1), 60-64.
- Mattila, A. S., & Wirtz, J. (2001). Congruency of scent and music as a driver of in-store evaluations and behavior. *Journal of Retailing*, 77(2), 273-289.
- Maxwell, K. A. (2002). Friends: The role of peer influence across adolescent risk behaviors. *Journal of Youth and Adolescence*, 31(4), 267-277.
- McElrea, H., & Standing, L. (1992). Fast music causes fast drinking. *Perceptual and Motor Skills*, 75(2), 362-362. doi: 10.2466/pms.1992.75.2.362
- North, A. C., & Hargreaves, D. J. (1997). Experimental aesthetics and everyday music listening *The social psychology of music*. (pp. 84-103). New York, NY US: Oxford University Press.
- North, A. C., & Hargreaves, D. J. (1999). Music and driving game performance. *Scandinavian Journal of Psychology*, 40(4), 285-292.
- Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 51(6), 768-774.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 37(1), 51-87.
- Rivers, S. E., Reyna, V. F., & Mills, B. (2008). Risk taking under the influence: A fuzzy-trace theory of emotion in adolescence. *Developmental Review*, 28(1), 107-144.
- Steinberg, L. (2008a). A dual systems model of adolescent risk-taking. *Developmental Psychobiology*, 52(3), 216-224.
- Steinberg, L. (2008b). A social neuroscience perspective on adolescent risk-taking. *Developmental Review*, 28(1), 78-106.
- van der Zwaag, M., Dijksterhuis, C., de Waard, D., Mulder, B., Westerink, J., & Brookhuis, K. (2011). The influence of music on mood and performance while driving. *Ergonomics*, 55(1), 12-22.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York, NY US: Cambridge University Press.
- Zuckerman, M., Kolin, E. A., Price, L., & Zoob, I. (1964). Development of a sensation-seeking scale. *Journal of Consulting Psychology*, 28(6), 477-482.

# Time your stress if you aim for success: the influence of stress on memory for different memory phases

## REVIEW

In the field of memory research, the influence of stress on memory is still unclear. The aim was therefore to review existing literature on the topic and to determine whether memory is facilitated or impaired by stress. A model is reviewed that hypothesizes that stress improves learning when it is experienced in the context and around the time of the stressful event. This is achieved via catecholamine and non-genomic glucocorticoid actions. Conceptually, the hormones shift the brain into a memory formation mode that facilitates encoding and suppresses the retrieval of irrelevant information. When catecholamine levels have returned to baseline, slow genomic glucocorticoid actions subsequently shift the brain into a memory storage mode where encoding of new irrelevant information will be suppressed to allow successful consolidation of information regarding the stressful event. It was concluded that stress can both enhance and impair memory depending on the timing of the stressful event.

**Keywords:** Stress, glucocorticoid actions, memory formation, memory storage.

Floor van den Brand; Bachelor student Psychology  
Maastricht University, Maastricht, the Netherlands

---

F.vandenbrand@student.maastrichtuniversity.nl

## INTRODUCTION

Stress has become a popular topic in our society, both judging by the articles that appear frequently in the media about its negative impact on human health, and by the many stress management therapies that are offered these days. Chronic stress is a risk for physical health problems: stress hormones such as glucocorticoids antagonize insulin and increase blood pressure, which increases the risk for developing diabetes and cardiovascular disease (Hjortskov et al., 2004; Kivimäki et al., 2002). Furthermore, stress impairs tissue growth and repair and suppresses immune functions which in turn increases the risk of infection (Chrousos, 2009). In addition, prolonged stress can lead to mental health problems such as burnout, which is mainly characterized by feelings of emotional exhaustion (Hooftman, 2011). Burnout is a major and still growing cause of work disability. In the Netherlands in 2010, about 13 percent of the working population experienced burn out problems (Hooftman, 2011).

Although it has become clear that (chronic) stress could have detrimental effects on human health, also positive effects of stress have been demonstrated in the field of memory research (Buchanan & Lovallo, 2001). An emotionally arousing experience can have an enhancing effect on memory. For example, when a boy is attacked by a dog in his childhood, it is likely that he still remembers this experience vividly years later. An extreme case of the enhancing effect of emotions on memory is a flashbulb memory. This is a highly vivid memory for an intense, emotionally engaging event (Hamann, 2001). Most people for instance will still remember vividly what they were doing on the day the Twin Towers were attacked. From an evolutionary perspective, the memory enhancing effect of emotional experiences is understandable since a highly emotionally arousing event is likely to have an effect on immediate or future survival, and therefore it will be important to be remembered for future occasions (Hamann, 2001). Even though the enhancing effects of acute stress on memory are clearly documented, stress or emotional arousal can also impair memory (Schwabe & Wolf, 2010). A typical example is when a person forgets a doctor's appointment as a consequence of being under a lot of pressure at work. In sum, research studies provide mixed results on the effect of stress on memory. Both enhancing and impairing effects of stress on learning are reported. It is still unclear how stress influences memory processes and what factors determine the performance outcome.

In sum, stress is a frequently occurring and increasing phenomenon in the working environment, therefore it is relevant to investigate to which degree experienced stress can be either harmful or beneficial. The aim of this paper is to review existing literature on this topic and to investigate which factors determine whether stress causes an enhancing or an inhibiting effect on memory. First, the nature of the stress response in the human body including involved brain areas and released stress hormones is explained. Second, the effects of stress on different memory stages are discussed while the influence of the valence of the learned material will be taken into account. Third, theories that can explain the opposing effects of stress on memory are described. Finally, this paper discusses possible effects of daily stressors on job performance and implications for everyday life.

## THE STRESS RESPONSE

When confronted with a stressful event, the body reacts by activating two biological systems: the rapidly acting autonomic nervous system (ANS) and the slower hypothalamic-pituitary-adrenal (HPA) axis (Schwabe, Wolf, & Oitzl, 2010). Important organs of the ANS are the adrenal glands, located above the kidneys. The inner portion of the adrenal glands, the adrenal medulla, releases a combination of hormones named catecholamines consisting of adrenaline and noradrenaline (Schwabe et al., 2010). These stress hormones act on the body to prepare an organism to a fight, flight or freeze response, for example by increasing heart rate and blood pressure (Lupien, Maheu, Tu, Fiocco, & Schramek, 2007). Adrenaline cannot enter the brain directly, but exerts its influence by activating the sensory vagus outside the brain and sending information via the nucleus of the solitary tract and the locus coeruleus into the brain. An important brain structure containing adrenergic receptors is the amygdala, which plays a role in fear processing and memory for emotionally relevant information (Adolphs, Tranel, & Damasio, 1998; Davis, 1992). The activation of the slower HPA system starts in the brain, in the paraventricular nucleus of the hypothalamus where corticotrophin releasing hormone (CRH) is secreted. CRH causes secretion of adrenocorticotrophic hormone (ACTH) from the anterior pituitary, a gland connected to and located beneath the hypothalamus. ACTH is then transported in the blood stream until it reaches the cortex of the adrenal glands where it induces the secretion of glucocorticoids (GCs).

In humans, the most abundant glucocorticoid is cortisol (called corticosterone in rodents). In response to a stressor, glucocorticoids trigger a chain of hormonal events that prepare the body for action, and lead for instance to an increase in heart rate and blood pressure. Glucocorticoids trigger different effects in the body that increase the availability of energy substrates in an organism and allow the organism to adapt to the changing environment and restore homeostasis. Since glucocorticoids are liposoluble, they can cross the blood-brain barrier to bind with receptors located in different parts of the brain. Three of the most important brain structures containing glucocorticoid receptors are the hippocampus, amygdala and frontal lobes; all areas known to be involved in learning and memory (Lupien et al., 2007).

## STRESS EFFECTS ON MEMORY: A MATTER OF TIMING?

Both memory enhancement (Cahill, Gorski, & Le, 2003) and impairment (Buchanan, Tranel, & Adolphs, 2006; Kirschbaum, Wolf, May, Wippich, & Hellhammer, 1996) due to stress have been reported. To explain how stress can lead to these opposing effects, a theory was proposed stating that the timing of the stressor relative to the memory phase determines the memory outcome (Joels, Pu, Wiegert, Oitzl, & Krugers, 2006; Roozendaal, 2002).

### **Stress effects on encoding**

Encoding is the first step in the process of creating a new memory. In one of the early human studies investigating the effect of stress on encoding, the effects of a psychological stressor as well as a physical stressor on learning were investigated (Kirschbaum et al., 1996). To induce psychosocial stress the Trier Social Stress Test (TSST) was used, which requires participants to deliver a speech in front of an audience. It reliably elicits moderate stress in laboratory settings and has shown to increase cortisol levels from  $8.46 \pm 1.02$  (SE) nmol/l before the test to  $17.65 \pm 2.17$  nmol/l after the stress induction (Kirschbaum et al., 1996).

In the second part of Kirschbaum's experiment, 10mg cortisol was administered to investigate the effect of physical stress independent of psychological stress. Participants had to learn a list of 26 nouns and recall the words after a short distraction task. The increase in cortisol level in the blood induced by the psychosocial stress task was inversely correlated with the number of remembered words from the list. The same memory-impairing effect was found for the group that received the synthetic cortisol. From this research, it can be concluded that stress and elevated cortisol levels has the potential to impair learning.

In a different study, participants underwent a cold pressure test (a test where participants are instructed to hold their hand in ice water as long as possible) ten minutes before positive, negative, and neutral words were presented (Schwabe, Bohringer, Chatterjee, & Schachinger, 2008). During the subsequent 1-h delayed free recall test, memory enhancement was found for the group that experienced pre-learning stress. The subjects who experienced a stress-induced cortisol elevation showed enhanced memory for negative terms but not for neutral words. The authors argue that this difference is due to the amygdala, which has a role in mediating the effects of stress on memory, and only processes emotionally valent material.

The facilitating effect of emotionally arousing material was confirmed in research that used a more complex memory task designed to resemble a real life situation instead of a simple word list (Payne et al., 2006). After stress induction using the TSST, the participants were shown a detailed slide show containing 9 neutral and 3 emotionally arousing slides combined with a narrative. This study found that stress led to an impaired memory for neutral information, while memory for emotional information was preserved (Payne et al., 2006).

From these studies, it is not entirely clear whether stress impairs or enhances encoding of memories. It seems that stress impairs memory for neutral material, but that it can protect or even enhance memory for emotionally arousing stimuli. However, it is not certain that these reported effects on memory are entirely attributable to the encoding phase. It is likely that stress induced shortly before a training session not only influences encoding during the learning phase, but also subsequent stabilization of the memory, i.e. memory consolidation.

Also, differences in design of the afore-mentioned studies may be responsible for diverging results. Furthermore, in the study of Kirschbaum et al. (1996), subjects were instructed to learn the word list, while in other research protocols participants were given instructions that kept them oblivious to the true purpose of the task. It is therefore possible that subjects in the Kirschbaum study used the delay period to rehearse the learning material which may have interfered with the results.

### **Stress effects on consolidation**

After the initial acquisition, a new memory trace is stabilized during the consolidation phase. The enhancing effects of stress on consolidation are mediated by the release of adrenal hormones like adrenaline and GCs that are released by emotional arousal. The brain contains two types of glucocorticoid receptors: the high-affinity mineralocorticoid receptors (MRs) and the low-affinity glucocorticoid receptors (GRs). A striking difference between the two receptor types is that the MRs bind glucocorticoids with an affinity that is 6-10 times higher than the GRs (Lupien et al., 2005; Reul & de Kloet, 1985). Under normal conditions, the glucocorticoid secretion has a 24-h circadian rhythm with a maximum concentration in the morning which slowly declines during the day until it reaches the circadian trough in the evening (Lupien et al., 2005).

It was hypothesized that activation of GRs and not MRs are involved in memory consolidation, since GRs become mainly occupied during stress when levels of glucocorticoids are high, while MRs are almost fully occupied under basal conditions (Reul & de Kloet, 1985; Roozendaal, 2000). Because it is therefore likely that GRs alone mediate the stress effects on memory consolidation, this was investigated by intracerebroventricular administration of specific antagonists of GRs or MRs in rats before or immediately after a training session in a Morris water maze task (De Kloet, Oitzl, & Joëls, 1999). In the experiment, only infusions of a GR antagonist impaired retention of the task 24h later. Treating the rats with the GR antagonist before the retrieval session was ineffective, which leads to the conclusion that GR inhibition interferes with the consolidation phase and not the retrieval of the learned spatial information (De Kloet et al., 1999). When the MRs were inhibited, the rats did not need more time to find the platform, but in a free swim trial their search pattern changed. They still swam directly to the former location of the platform, which shows their retention was still intact, but different from the controls that remained searching in the vicinity of the platform, they subsequently explored other areas of the pool. This distinct behavior was explained by the authors as MR activation induced increase in behavioral reactivity (approaching and investigating a stimulus) which depends on the hippocampus (De Kloet et al., 1999). A similar increase in reactivity was found in the absence of corticosteroids after adrenalectomy and with a high occupation of both MR and GR following high doses of corticosterone. This line of research confirms the hypothesis that GRs are involved in regulating glucocorticoid effects on memory consolidation, while MR activation is important for the interpretation of environmental stimuli and is involved in the selection of a behavioral response (De Kloet et al., 1999).

To explain the effects of glucocorticoids on cognitive performance, the MR/GR ratio hypothesis was developed, which suggests that the proportion of occupied GR and MR receptors rather than separate receptor activation determines memory performance. This can be described as an inverted u-shaped function (De Kloet et al., 1999). When most of the MR and only a part of the GR receptors are activated, cognitive functioning is optimal (top of the curve). However, when the MR/GR ratio is low as a result of a decrease or increase of circulating glucocorticoid levels (both extremes of the function), memory performance will be impaired.

The MR/GR ratio hypothesis was investigated in humans using a hormone

removal-replacement protocol (Lupien et al., 2002). In the experiment of Lupien et al. (2002), glucocorticoid levels were pharmacologically lowered using metyrapone, a glucocorticoid synthesis inhibitor, and subsequently restored by infusing hydrocortisone, a synthetic glucocorticoid. The results showed that when circulating glucocorticoid levels were lowered, memory performance significantly decreased. Memory impairment was completely reversed when glucocorticoid levels were restored after hydrocortisone replacement. This experiment proves that the absence of glucocorticoids impairs memory.

A few years earlier, an experiment had been performed to investigate the influence of high circulating levels of GCs on memory, which would match the right extreme of the function (Lupien, Gillin, & Hauger, 1999). The researchers infused a dose of glucocorticoids in the morning at the time of the circadian peak to create a very high glucocorticoid concentration, and indeed found a severely impaired memory function. This research shows that the time of testing should be chosen carefully since it could influence the outcome of the experiment. Applying a stressor in the morning can impair memory performance while the same stressor experienced in the afternoon could increase memory function (Lupien et al., 2005). Together, these studies show that glucocorticoids can have different behavioral effects. This can be attributed to the distinct functions of the MRs and GRs and the proportion of receptor occupation as explained by the MR/GR ratio hypothesis. As for the influence of stress on memory: stress during the consolidation phase seems to enhance memory by stimulating the formation of a stable memory trace. The memory enhancing effects are attributed to the activation of the glucocorticoid receptors by the stress hormone cortisol. There is evidence that this memory enhancement applies especially to emotionally arousing stimuli due to a mediating role of adrenal hormones.

### **Stress effects on retrieval**

Retrieval refers to the process of accessing information that has been saved in memory. Several studies investigated the effect of stress on retrieval (Buchanan et al., 2006; De Quervain et al., 2003; De Quervain, Roozendaal, Nitsch, McGaugh, & Hock, 2000; Kuhlmann, Kirschbaum, & Wolf, 2005; Tollenaar, Elzinga, Spinhoven, & Everaerd, 2009).

The impairing effect of cortisol on delayed free recall was demonstrated in a research using only neutral words (De Quervain et al., 2000). In a different study, both neutral and negative words were used to assess the effect of the valence of the material on memory (Kuhlmann, Kirschbaum, et al., 2005). The effect of cortisol administration was measured 5h after an intentional learning session. Performance on free recall, cued recall and working memory (using a digit span task) was compared between a cortisol and placebo group. Cortisol significantly impaired retrieval of negative words while having no significant effect on neutral words. Another research using both emotional and neutral words used a delay period of an entire week before assessing retrieval performance. This study demonstrated that cortisol administration before recall lead to a significant decrease in memory performance compared to a placebo group. This memory impairment caused by cortisol remained even when recall was tested again after a washout period of one



week. Furthermore, emotional words were recalled better compared to neutral words, but no interaction with treatment group was demonstrated (Tollenaar et al., 2009).

These studies successively show that (i) cortisol can impair recall of (neutral) material, (ii) cortisol impairs memory but only for negative words and (iii) cortisol leads to a decrease in memory performance but that emotional material is remembered better than neutral items. These contradicting findings could be a consequence of different research designs, but mainly show that research is still inconclusive about the effects of stress on retrieval and possible interacting effects of valence of the learning material.

Besides valence of the learned material, research has focused on revealing what type of memory is affected by cortisol. Cortisol treatment has shown to impair free delayed recall, while cued recall and recognition memory were preserved (Kuhlmann, Kirschbaum, et al., 2005; Wolf et al., 2001). However, different research has shown impaired cued recall as well (De Quervain et al., 2003). While the type of recall method leads to different results, researchers agree that cortisol seems to influence declarative memory specifically, and does not affect working memory, attention, verbal executive function or vigilance (De Quervain et al., 2000; Kuhlmann, Kirschbaum, et al., 2005; Tollenaar et al., 2009; Wolf et al., 2001).

It was hypothesized that the hippocampus mediates the negative effects of cortisol on recall (Wolf et al., 2001), since it has been demonstrated that hippocampal activation is important for successful memory retrieval (Schacter & Wagner, 1999). In order to investigate this theory, a PET study was conducted which tested the 24h delayed recall of a word list (De Quervain et al., 2003). Cortisone administration before recall significantly impaired cued recall and induced a decrease in regional cerebral blood flow in the right posterior medial temporal lobe, with a maximal decrease in the parahippocampal gyrus. This latter brain area is associated with successful verbal memory retrieval (Schacter & Wagner, 1999), which suggests that elevated glucocorticoid levels can disturb medial temporal lobe function and thereby impair declarative memory retrieval (De Quervain et al., 2003). This result is in line with earlier research that showed a cortisol-induced reduction in hippocampal glucose metabolism (De Leon et al., 1997).

## THEORIES ON THE DIFFERENTIAL EFFECT OF STRESS ON MEMORY

The studies discussed in the previous section have shown that stress can have enhancing as well as detrimental effects on memory, depending on the memory phase during which the stress was experienced. Different theories have been developed in an attempt to explain this phenomenon.

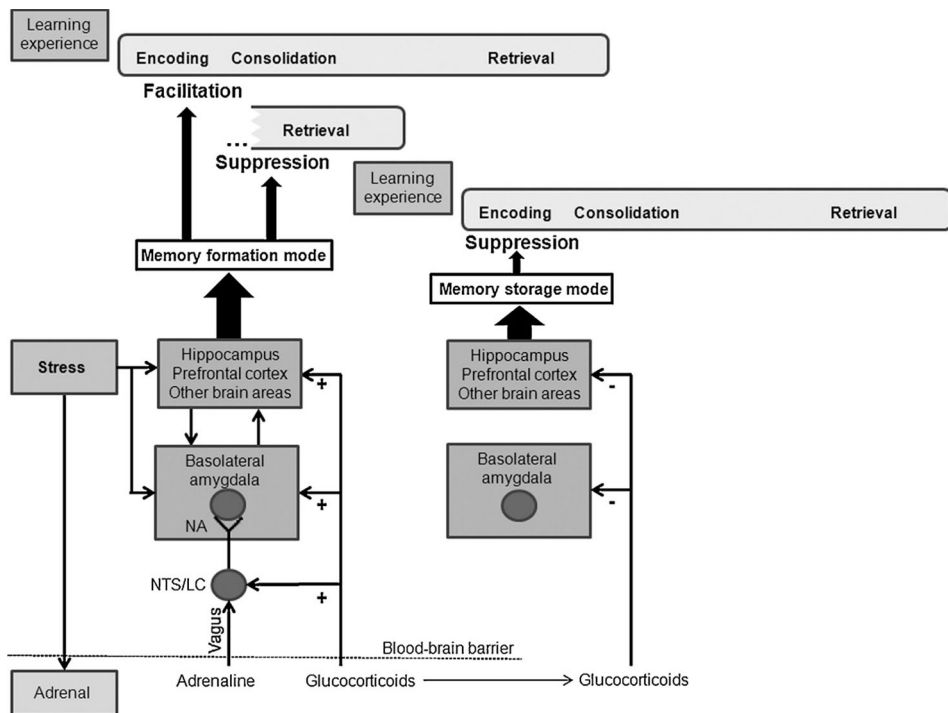
In order to explain the seemingly paradoxical time-dependent effects of stress on memory, Joëls et al. (2006) state that learning will only be facilitated when stress is experienced in the context and around the time of the event that needs to be remembered. Stress will enhance memory when neurotransmitters and hormones that are released in reaction to a stressful situation act upon the same neuronal circuits that are involved in the processing of the information (convergence in



space). Also, the increase of circulating stress hormones has to take place at about the same time that these circuits are activated by the event (convergence in time), and the stress has to be experienced within the learning context (convergence in context).

### Memory formation mode model

Schwabe et al. (2012) combined and expanded the model of Joëls et al. (2006) with an earlier developed model that described stress-hormone induced interaction of brain systems (McGaugh, 2000) to form an integrated model (fig. 1). A stressful event stimulates the secretion of adrenaline and glucocorticoids. The rapid adrenaline and non-genomic glucocorticoid effects interact in the basolateral amygdala, which causes other brain areas including the hippocampus and the prefrontal cortex to shift into a 'memory formation mode'. In this memory formation mode the processing of the stressful event will be stimulated by facilitating the encoding of information about the situation and environment and by suppressing other cognitive processes, which might interfere with the consolidation process. Over time, the slower genomic glucocorticoid actions become active leading to a 'memory storage mode'. In this mode, encoding of new information will be suppressed to reduce interference with memory consolidation. Possibly, the retrieval of old information will also be impaired. By suppressing encoding of new material, the storage mode promotes consolidation and long-term storage of relevant information about the stressful experience (Schwabe, Joels, Roozendaal, Wolf, & Oitzl, 2012).



**Figure 1.** Memory formation mode model of stress effects on memory. NA: noradrenaline; NTS: nucleus tractus solitarius; LC: locus coeruleus. Reprinted with permission from Schwabe et al. (2012).

## DISCUSSION

The aim of this paper was to investigate the effects of stress on memory. A large body of research studies suggests that stress has opposing effects for each of the memory stages, because timing determines whether stress enhances or impairs memory. The differential effects of stress can be explained by the immediate actions of catecholamines and non-genomic glucocorticoids that shift the brain into a memory formation mode, and the slower genomic glucocorticoid actions that cause a memory storage mode. These different brain modes facilitate encoding and consolidation of relevant information about the stressful event and, at the same time, increase the threshold for information unrelated to the stressor. This system ensures efficient information processing and storage of threatening events, which may be important for future survival. The different effects of stress on memory performance can thus partially be explained by differences in timing of stress relative to the training phase. However, when studies are categorized by timing of the stressor, results are still incomprehensive.

The MR/GR ratio hypothesis could further clarify these opposing findings. As suggested by the inverted u-shaped function, the proportion mineralocorticoid and glucocorticoid receptors rather than the separate receptor activation determines memory performance. When most of the MR and only a part of the GR receptors are activated, cognitive functioning is optimal but if the MR/GR ratio is low as a result of a decrease or increase of circulating glucocorticoid levels memory performance will be impaired.

Since human circulating glucocorticoid levels follow a 24-h circadian rhythm with a peak in the morning and a trough in the afternoon, the time of day at which the experiment is carried out is essential for the results. Glucocorticoid intake when internal levels are low could improve memory while a dose of cortisol when circulating hormone levels are high could be detrimental for learning. Additionally, the ingested dosage could make the difference between memory impairment and stimulation. Every study used different doses of glucocorticoids, which could partially explain differing results, since it has been shown that only optimal glucocorticoid levels in the blood enhance performance while higher or lower levels could impair learning and memory (Lupien et al., 1999; Lupien et al., 2002).

In most studies on the influence of stress on memory a single event was used to invoke a short-lasting stress response in the participants. It is therefore still unknown how (chronic) daily life stressors affect memory and whether, for example, work stress would have impairing or stimulating effects on job performance. Based on the before discussed studies (De Quervain et al., 2000; Kuhlmann, Piel, & Wolf, 2005; Tollenaar et al., 2009), it seems probable that under stressful circumstances like examinations or approaching deadlines, elevated stress hormones could induce problems with remembering information. On the other hand, experiencing stress could cause a person to shift his cognitive resources to the important situation in question for efficient information processing and consolidation without distraction from irrelevant stimuli. Furthermore, prolonged exposure to stress may induce changes in hormone and neurotransmitter systems and could cause functional and structural changes in the hippocampus that would lead to impaired memory

storage (De Quervain et al., 2000).

Intuitively, it seems likely that trying to remember information is harder when feeling stressed or aroused. However, participants usually report no side effects or strange feelings after cortisol treatment (De Quervain et al., 2003; Tollenaar et al., 2009). Although participants thus seem unaware of the artificially induced stress, research clearly shows impairing effects of stress on memory performance. This could mean that in daily life, elevated stress levels could go unnoticed by an individual, while still having detrimental effects on his memory and performance.

Finally, it can be stated that stress not only influences how much is learned, but also affects how information is learned and what aspects of the information will be remembered. In other words, besides the quantity, also the quality of memory is influenced by stress. Stress could influence which (mal)adaptive learning strategies people use in their daily lives, and can affect how people behave in response to work-related dilemma's (Schwabe et al., 2007).

To conclude, the literature shows that stress can have both a stimulating and impairing influence on memory, depending on the context and timing of the stressful event. Whether these stress effects are similar in daily life situations still needs to be determined.

## REFERENCES

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, 393(6684), 470-474.
- Buchanan, T. W., & Lovaglio, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, 26(3), 307-317.
- Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning & memory*, 13(3), 382-387.
- Cahill, L., Gorski, L., & Le, K. (2003). Enhanced human memory consolidation with post-learning stress: interaction with the degree of arousal at encoding. *Learning & memory*, 10(4), 270-274.
- Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nat Rev Endocrinol*, 5(7), 374-381.
- Davis, M. (1992). The role of the amygdala in fear and anxiety. *Annual review of neuroscience*, 15, 353-375.
- De Kloet, Oitzl, M. S., & Joëls, M. (1999). Stress and cognition: are corticosteroids good or bad guys? *Trends in Neurosciences*, 22(10), 422-426.
- De Leon, M. J., McRae, T., Rusinek, H., Convit, A., De Santi, S., Tarshish, C., Golomb, J., Volkow, N., Daisley, K., Orentreich, N., & McEwen, B. (1997). Cortisol reduces hippocampal glucose metabolism in normal elderly, but not in Alzheimer's disease. *The Journal of clinical endocrinology and metabolism*, 82(10), 3251-3259.
- De Quervain, D. J., Henke, K., Aerni, A., Treyer, V., McGaugh, J. L., Berthold, T., Nitsch, R. M., Buck, A., Roozendaal, B., & Hock, C. (2003). Glucocorticoid-induced impairment of declarative memory retrieval is associated with reduced blood flow in the medial temporal lobe. *The European journal of neuroscience*, 17(6), 1296-1302.
- De Quervain, D. J., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans.

- Nature neuroscience*, 3(4), 313-314.
- Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences*, 5(9), 394-400.
- Hjortskov, N., Rissén, D., Blangsted, A., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92(1-2), 84-89.
- Hoofman, W., Koppes, L.L.J., Vroome, E.M.M. de., Kraan, K., Driessen, M., & Bossche, S.N.J. van den. (2011). NEA 2010 : vinger aan de pols van werkend Nederland. Hoofddorp: TNO.
- Joels, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: how does it work? *Trends in cognitive sciences*, 10(4), 152-158.
- Kirschbaum, C., Wolf, O. T., May, M., Wippich, W., & Hellhammer, D. H. (1996). Stress- and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life sciences*, 58(17), 1475-1483.
- Kivimäki, M., Leino-Arjas, P., Luukkonen, R., Riihimäi, H., Vahtera, J., & Kirjonen, J. (2002). Work stress and risk of cardiovascular mortality: prospective cohort study of industrial employees. *BMJ*, 325(7369), 857.
- Kuhlmann, S., Kirschbaum, C., & Wolf, O. T. (2005). Effects of oral cortisol treatment in healthy young women on memory retrieval of negative and neutral words. *Neurobiology of learning and memory*, 83(2), 158-162.
- Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(11), 2977-2982.
- Lupien, S. J., Fiocco, A., Wan, N., Maheu, F., Lord, C., Schramek, T., & Tu, M. T. (2005). Stress hormones and human memory function across the lifespan. *Psychoneuroendocrinology*, 30(3), 225-242.
- Lupien, S. J., Gillin, C. J., & Hauger, R. L. (1999). Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: a dose-response study in humans. *Behavioral neuroscience*, 113(3), 420-430.
- Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and cognition*, 65(3), 209-237.
- Lupien, S. J., Wilkinson, C. W., Briere, S., Menard, C., Ng Ying Kin, N. M., & Nair, N. P. (2002). The modulatory effects of corticosteroids on cognition: studies in young human populations. *Psychoneuroendocrinology*, 27(3), 401-416.
- McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, 287(5451), 248-251.
- Payne, J. D., Jackson, E. D., Ryan, L., Hoscheidt, S., Jacobs, J. W., & Nadel, L. (2006). The impact of stress on neutral and emotional aspects of episodic memory. *Memory*, 14(1), 1-16.
- Reul, J. M., & de Kloet, E. R. (1985). Two receptor systems for corticosterone in rat brain: microdistribution and differential occupation. *Endocrinology*, 117(6), 2505-2511.
- Roozendaal, B. (2000). 1999 Curt P. Richter award. Glucocorticoids and the regulation of memory consolidation. *Psychoneuroendocrinology*, 25(3), 213-238.
- Roozendaal, B. (2002). Stress and memory: opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of learning and memory*, 78(3), 578-595.
- Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus*, 9(1), 7-24.
- Schwabe, L., Bohringer, A., Chatterjee, M., & Schachinger, H. (2008). Effects of pre-learning stress on memory for neutral, positive and negative words: Different roles of cortisol and autonomic arousal. *Neurobiology of learning and memory*, 90(1), 44-53.
- Schwabe, L., Joels, M., Roozendaal, B., Wolf, O. T., & Oitzl, M. S. (2012). Stress effects on memory: an update and integration. *Neuroscience and biobehavioral reviews*, 36(7), 1740-1749.
- Schwabe, L., Oitzl, M. S., Philippsen, C., Richter, S., Bohringer, A., Wippich, W., & Schachinger,

- H. (2007). Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learning & memory*, 14(1), 109-116.
- Schwabe, L., & Wolf, O. T. (2010). Learning under stress impairs memory formation. *Neurobiology of learning and memory*, 93(2), 183-188.
- Schwabe, L., Wolf, O. T., & Oitzl, M. S. (2010). Memory formation under stress: quantity and quality. *Neuroscience and biobehavioral reviews*, 34(4), 584-591.
- Tollenaar, M. S., Elzinga, B. M., Spinhoven, P., & Everaerd, W. (2009). Immediate and prolonged effects of cortisol, but not propranolol, on memory retrieval in healthy young men. *Neurobiology of learning and memory*, 91(1), 23-31.
- Wolf, O. T., Convit, A., McHugh, P. F., Kandil, E., Thorn, E. L., De Santi, S., McEwen, B. S., & de Leon, M. J. (2001). Cortisol differentially affects memory in young and elderly men. *Behavioral neuroscience*, 115(5), 1002-1011.

# Fear to get near: personal space in individuals with psychopathic traits

## ORIGINAL PAPER

Past literature theorized that personal space -a physical boundary in which a person experiences discomfort when another person intrudes it- might depend on personality. This study hypothesized, that higher scores on psychopathic personality measures would lead to a larger personal space in individuals. Furthermore, as personal space is thought to function as an intraspecies aggression regulator, this study also hypothesized that personal space tends to be larger towards dominant individuals, than towards non-dominant individuals, as dominant other might be seen as possibly more aggressive. Results indicate no direct correlation between overall psychopathic traits and personal space. However, a positive correlation is found in the relationship between one particular psychopathic personality trait -Coldheartedness- and personal space. Regarding perceived dominance, it was found that individuals tested in this study kept a larger personal space towards dominant individuals. This effect was found to be independent of score on a psychopathic personality measure.

**Keywords:** psychopathy, personal space, perceived dominance

Martijn van Teffelen; bachelor student year 3  
Maastricht University, Maastricht, the Netherlands

---

[m.vanteffelen@student.maastrichtuniversity.nl](mailto:m.vanteffelen@student.maastrichtuniversity.nl)

## INTRODUCTION

### **Psychopathy**

The psychopathic personality is a concept that was first systematically observed and reported in “The Mask of Sanity” by Hervey Cleckley (1941). Based on this work, Cleckley outlined 16 criteria for the assessment of the psychopathic personality, or psychopathy, such as unreliability, untruthfulness and insincerity, lack of remorse or shame, inadequately motivated antisocial behavior, pathologic egocentricity and incapacity for love (Cleckley, 1941). These criteria served as the basis for the so called Psychopathy Checklist (Hare, 1980), and later the Psychopathy Checklist-Revised (PCL-R) (Hare, 1991), a clinical, observational assessment of psychopathy. The factors ‘selfish, callous and remorseless use of others’, and ‘chronically unstable, antisocial and socially deviant lifestyle’ in this measurement were derived through principal components analysis from Cleckley’s (1941) original criteria (Hare, 1991). In addition, Blackburn (1975) classified two types of psychopathy using self-reports: primary and secondary psychopathy. Primary psychopaths tend to be extroverted, self-confident and low-anxious, whereas secondary psychopaths are characterized by social anxiety, moodiness, low self-esteem and social withdrawal (Blackburn, 1975; Morrison & Gilbert, 2001).

Furthermore, it should be noted that there are explicit gender differences in the general population, when men and women are measured with the Psychopathy Personality Inventory-Revised (PPI-R), a self-report psychopathy questionnaire. In this, men tend to score significantly higher than women (Lilienfeld & Andrews, 1996).

### **History of psychopathic measures**

In their paper, Lilienfeld and Andrews (1996) outlined two points of criticism on the PCL-R, firstly concerning the absence of items assessing anxiety. Therefore, it differs from Cleckley’s (1941) original concept of the psychopathic personality (Blackburn, 1975; Lilienfeld & Andrews, 1996; Morrison & Gilbert, 2001). In addition, the PCL-R is only suited to measure psychopathy in the clinical population, since the PCL-R was developed as a clinical observation scale. This does not do right to the fact that psychopathy was originally outlined by Cleckley (1941) as a personality trait, pleading for a more dimensional approach to the concept. On this account, the Psychopathy Personality Inventory (PPI, 181 items) (Lilienfeld & Andrews, 1996), and later the Psychopathy Personality Inventory-Revised (PPI-R, 154 items) (Lilienfeld & Widows, 2005) self-report measures were developed.

The PPI has two advantages compared to the PCL-R. Firstly, these measures are based on a dimensional approach as psychopathy is seen as a continuous variable. Thus it assumes that an individual can possess psychopathic personality traits to a larger or smaller extent. Secondly, the PPI contains items measuring anxiety. The PPI-R contains the subscales Stress Immunity, Fearlessness, Social Potency (e.g. dominance), Carefree Nonplanfulness (impulsivity), Rebellious Nonconformity, Blame Externalization, Machiavellian Egocentricity (aggressive exploitation) and Coldheartedness (lack of empathic concern). These subscales served to be the



variables of interest that loaded on the following three factors by derivation through principal component analysis: PPI-I, characterized as Fearless Dominance, is loaded on by Fearlessness, Social Influence and Stress Immunity; PPI-II, characterized as Self-Centered Impulsivity, is loaded on by Carefree Nonplanfulness, Rebellious Nonconformity, Blame Externalization and Machiavellian Egocentricity; and The third factor, PPI-III is loaded on by Coldheartedness alone (Lilienfeld & Widows, 2005).

### **Personal space and aggression**

It is assumed that every human being posits a personal space (PS), or a physical boundary in which people experience discomfort when another person intrudes it (Sommer, 1959). This would mean that individuals with a large PS tend to keep larger physical distances between themselves and others, compared to individuals with a small PS. It is proposed that this area functions to reduce intraspecies aggression, as intrusion of this space could lead to aggression (Pfeiffer, 1969). According to that, it has been suggested that personal space invasions of individuals with a larger personal space and a tendency to react with fight rather than flight – a psychophysiological response towards threatening stimuli-, might result into assaultive behavior (Eastwood, 1985). This is supported by clinical observations of psychopathologic individuals. For example, PS is supposedly greater in schizophrenics (Sommer, 1959), but also in violent offenders (Kinzel, 1970; Newman & Pollack, 1973; Wormith, 1984) - both prone to assaultive behavior. In practice this means that these individuals have a reduced tolerance towards individuals who come physically close to them.

A study by Eastwood (1985) contradicts these findings, as it did not find evidence for correlation between PS and violent behavior. However, the study found an interaction-effect between violent behavior, PS and psychoticism – referring to aggressiveness and interpersonal hostility, found in one's personality (Eysenck & Eysenck, 1975). This suggests an individual's PS might not only be dependent on one's violent nature, but also on certain personality traits.

Research on the topic is rather limited. Preliminary studies in prison populations by Kinzel (1970) and Newman and Pollack (1973) did not mention personality, nor it's possible correlation with PS. A study by Wormith (1984) reported mild negative correlations between personal space and empathy, ( $r = -.37, p < .01$ ), and acceptance of others ( $r = -.33, p < .05$ ). Wormith concluded from this finding, that violent offenders with a large PS seem to be less empathic and less acceptant to others.

### **Personal Space and Dominance**

Recent studies show renewed interest in studying PS in relation to a specific personality trait, called dominance (Hall, Coats, & LeBeau, 2005), as the amount of research in the preceding years seems limited on search engines such as PsycInfo and PubMed. A meta-analysis by Hall et al. (2005) suggests that PS is negatively correlated to dominance, with correlation coefficients ranging from  $-.08$  to  $-.22$ . This suggests that individuals with a high level of dominance tend to have a small PS. Although psychopathic traits in this meta-analysis were not discussed, it seems

to contrast findings seen in psychopathic individuals, who are presumed to be dominant and tend to have a large PS (Kinzel, 1970; Lilienfeld & Widows, 2005; Newman & Pollack, 1973).

However, it is currently unknown in which way an individual's PS relates to the perceived dominance of others, and how psychopathic traits might have an influence on this. When PS indeed regulates intraspecies aggression, it should then be expected that when a highly dominant individual approaches, one's PS enlarges, as dominant others might be seen as an aggressive threat.

### The present study

Past research suggests that a possible interaction-effect might exist between PS, violence, and aggressive personality traits (Eastwood, 1985; Kinzel, 1970; Newman & Pollack, 1973; Wormith, 1984). However, support for this claim is limited up to this date. In addition, observations were made mainly in clinical populations, while more recent research suggests that aggressive personality traits are continuous rather than dichotomous, and should therefore also be present in nonclinical populations. As aggressive personality traits such as psychopathy are presumably of continuous nature, and these traits are possibly linked to personal space, it is reasonable that a relationship might exist between these traits and PS in both clinical and non-clinical populations. The present study therefore hypothesizes that psychopathic personality traits in healthy male adults are positively correlated with their PS. Subsequently, this study hypothesizes that PS is larger in healthy male adults when a dominant male adult approaches, than when a non-dominant, or even submissive male adult approaches.

**Table 1:** Participant demographics

<b>N</b>	70
<b>Mean age (SD)</b>	29.67 (13.56)
<b>Occupation</b>	
Unemployed	0
Employed	30
Student	39
Retired	1
<b>Highest completed education</b>	
primary school	0
high school	28
college/university	42
<b>Native language</b>	
Dutch	68
German	2

## METHODS

### Participants

Participants selected for this study were adult men from the non-clinical population (age 18-65), participating for either a monetary reward of €15,-, or course credit if they were students. They were approached either through poster-distribution at Maastricht University, or by receiving a phone call if they had registered themselves in a participant database.

In total, 93 participants participated in the experiment. The first one participated in test trials of the experiment, and was not included in the final analyses. Based on the Chauvenet's Criterion to detect outliers, two other participants were excluded (Chauvenet, 1960). Furthermore, eight participants were excluded from the data, as they did not pass the manipulation check – discussed in the materials section. Another eleven indicated during the manipulation check that they doubted the manipulation at some point during the experiment. For this reason, analyses were carried out without these cases on a total of 70 data sets. Table 1 shows participant demographics.

### Materials

#### *Psychopathy Personality Inventory Revised*

To assess psychopathic personality traits, the PPI-R (Lilienfeld & Widows, 2005) was used. This self-report questionnaire consists of 187 items, with three underlying factors – Fearless Dominance, Self-Centered Impulsivity, and Coldheartedness. Methodologically, satisfactory internal (PPI-R Total,  $\alpha = .91$ ; PPI-I,  $\alpha = .91$ ; PPI-II,  $\alpha = .89$ ; PPI-III,  $\alpha = .79$ ), construct (correlates ranging from  $r = .18$ -.68 with other psychopathy measures), and external validity has been reported for PPI-R factors (Uzieblo, Verschuere, Van den Bussche, & Crombez, 2010). Additionally, research by Sandler (2007) investigated the test-retest reliability between PPI-R factors and PPI-R total scores. Sandler reported the test-retest reliability for each factor and for PPI-R total score: PPI-total,  $r = .93$ ; PPI-R I,  $r = .91$ ; PPI-R II,  $r = .90$  and PPI-R III,  $r = .76$ )

#### *Stop-distance method*

PS is used as the main outcome measure in this study. To measure participants' PS, we subjected them to the so-called "stop-distance procedure" by Newman and Pollack (1973). This test was derived from the so-called 'body buffer zone- test' by Kinzel (1970), and was originally designed to measure the PS construct in violent offenders, as discussed in the introduction section. The PS can be commonly depicted in an ellipsoid way, representing mean distances of all directions, derived from the following formula (Wormith, 1984):

$$\text{Personal space} = \frac{\pi (F + B + 1) (R + L + 1)}{4}$$

In this formula F represents proxemic distance front, B represents proxemic distance behind, R represents proxemic distance right and L represents proxemic distance left. Subsequently the mean distance per direction (front, behind, left and right), and PS score are calculated, and used as outcome measures. Proxemic distances are defined as the distance between the individual that is being approached, and the approaching other individual.

### *Job-interest interview*

To manipulate perceived dominance, a job-interest interview took place with a dominant confederate, and a non-dominant confederate. The job-interest interview had the following structure for each participant. Participants were asked to pick one of three job preferences: 1) a non-interactive job (i.e. truck driver on long distances, night-time security guard), 2) a supervising job (i.e. head of a department), and 3) a non-supervising social job (i.e. bank employee). Then they were asked to explain why they had chosen this option, and what would make them qualified for this job, or why they would not be qualified. After this, nine hypothetical questions were asked, on what they would do if they were a supervisor/employee, independent of their initial choice -(i.e. “what would you do if you were a supervisor and you had an insecure employee who underestimated his capacities). This interview was developed for this study. To test the hypothesis whether dominance affects personal space, participants were exposed to two male confederates, confederate 1 acting submissive, and confederate 2 acting dominant. Confederates acted by keeping a closed, insecure attitude. He acted as if this was his first time leading a test and that he was insecure of how to act and what to do, talking soft, and literally reading everything from a provided script. Confederates acted dominant by keeping an open, extraverted attitude, overruling confederate 1 during interaction. In interacting with the participant, confederate 2 talked loudly, looked the participant straight in the eyes and pro-actively followed and maintained the conversation (e.g. by often nodding or humming when the participant spoke). Confederates were introduced as student, while confederate 2 was introduced as supervisor. The role of confederate 2 was carried out by three different fluently Dutch speaking adult men. The role of confederate 1 was carried out by one of the fluently actors throughout the experiment.

### **Procedure**

Since testing the hypotheses required both psychopathy assessment and dominance induction, a cover story was needed to avoid socially desired answering. Subjects were told they were participating in a study about personality and job interest. For this purpose, job interest was assessed by conducting a job-interest interview, discussed in the materials section.

Participants were welcomed to the study by confederate 1, and were asked to fill in the Dutch version of the PPI-R on a computer. After this, participants were

guided to an empty room, and PS was measured by confederate 1. The room was marked by a 2x2 square, containing a 30 centimeter grid. In the center of this square, a 10x10 centimeter square was marked, and this is where the participant stood during the test. The participants were explained that they would be approached from four different directions and that they should indicate the point at where they were starting to feel uncomfortable. They were supposed to keep their gaze directed straight forward. The experimenter approached the participant from the outside of the outer square (2 meters) by steps of 15 centimeters with his gaze directed to the ground. As they looked up, they waited three seconds for the participant to signal, and stopped as soon as he or she did. Each direction had two trials, in sequence front, left, behind, right, right, behind, left, front.

After this, participants were guided back to the lab, where they met confederate 2. At this point in time, confederate 2 started an a priori set up conversation with confederate 1, where he told him in a dominant way that the interview had gone wrong (e.g. "How can this still go wrong after testing so many participants?!"). Then confederate 2 proposed to the participants to redo the interview. If they conformed –which all included participants did–, the job interview took place again, only now led by dominant confederate 2. Afterwards, confederate 2 led the participants to the empty room, where their PSs were measured a second time. Finally, a manipulation check was carried out by asking the participants to fill in a response questionnaire.

### **Manipulation Check**

The manipulation check consisted of one item -with three possibilities to answer- assessing whether participants noticed that the experiment was staged (yes/no/doubt), and four items to assess on a visual analogue scale what participants thought of each of the different interviewers 1) "How nice did the submissive/dominant interviewer appear?", 2) How skillful did the submissive/dominant interviewer appear?, 3) "To what extent would you like this interviewer as your boss?", and 4) "To what extent would you like this interviewer as your employee?"

### **Statistical analyses**

Statistical analyses were performed using IBM SPSS 18 statistical software package. As indicated above, two hypotheses were to be tested. The first hypothesis concerned the correlation of psychopathic personality traits with personal space. To this extent, a positive correlation is expected between PPI-R score, or PPI-R factor scores and PS score. Therefore, larger PPI-R and PPI-R factor scores should result in larger PS scores. Two sets of regression analyses were performed using proxemic scores (front, behind, left and right distances and PS score) as outcome measures. The first set was carried out using PPI-R total score as single predictor. The second set used PPI-R factor scores -I, II and III- as predictors. To this extent, variables were initially entered and eliminated backwards.

Hypothesis two was that PS would be larger when approached by a dominant male adult, than when approached by a non-dominant male adult. In both hypotheses, all proxemic scores (front, behind, left, right and PS score) were used as dependent variables. The reason to use other proxemic scores in addition to PS

score as outcome measures was that the PS score proposed by Wormith (1984) is mathematically built up out of each of the separate proxemic distances. As can be seen from the formula, each proxemic distance carries the same mathematical weight. Nevertheless we found it important also to reveal the contribution of each proxemic distance to the overall PS score. It was proposed by Kinzel (1970) for example, that front and behind distance might be influenced by personality traits, whereas left and right distance might not be. The hypothesis that there may be such differences was tested by performing a paired-samples t-test on proxemic score obtained by non-dominant confederate 1 and dominant confederate 2.

Furthermore, to check whether the experiment was confounded, it was checked if proxemic scores were affected by individual differences in confederate 2 actors. To this extent, a one way ANOVA was carried out with proxemic scores (front, behind, left, right distance and PS score) as dependent variable and confederate 2-actor (one, two or three) as independent variable.

## RESULTS

### **Dominant confederates**

A one way ANOVA revealed that proxemic scores differed significantly between the different people who acted as confederate 2. Front distance ( $F(2,78) = 6.68$ ,  $p = .002$ ), behind distance ( $F(2,78) = 3.66$ ,  $p = .03$ ) as well as PS score ( $F(2,78) = 3.61$ ,  $p = .03$ ) differed significantly between actors. The scores for left distance ( $F(2,78) = 2.35$ ,  $p = .10$ ) and right distance ( $F(2,78) = 2.45$ ,  $p = .10$ ) were not significantly influenced by the different people who acted as confederate 2. One person that acted as confederate 2 in particular tended to get higher scores than the other two, meaning that he could approach the participants less closely.

In addition, a one way ANOVA was performed with dominant interviewer as independent variable, and score on manipulation check items as dependent variables (see methods section, for questions 1-4 regarding the dominant interviewer). This resulted in non-significant results for each question (Question one:  $F(2,88) = 1.50$ ,  $p = .23$ ; Question two:  $F(2,88) = 1.48$ ,  $p = .24$ ; Question three:  $F(2,88) = 1.50$ ,  $p = .23$ ; Question four:  $F(2,88) = 1.53$ ,  $p = .22$ ), meaning that the different people who acted as confederate 2 had no influence on how the questions were overall answered by the participants.

Since a significant effect of confederate 2-actor on proxemic scores was found, this has to be taken into account in the analyses regarding PS. Consequently, two regression analyses were run. One for the participants that were tested by the actors who did not lead to significant results and one for the participants that were tested by the one who lead to significant results. As a direct result statistical power dropped from .96 at  $N=81$  (analyses including SDC cases), to .86 at  $N=51$  (analyses excluding SDC cases).

### **Psychopathic traits and Personal Space**

The first hypothesis was that PPI-R total score and PPI-R factor scores possibly correlate positively with proxemic scores. Results are reported in table 2. Regression analyses performed excluding the SDC with PPI-R total score as single predictor, yielded a significant result. In this case, behind distance correlated positively with PPI-R total score ( $\beta = .30$ ,  $t(49) = 2.19$ ,  $p < .05$ ) in the dominant condition. This would mean that individuals with larger PPI-R total scores show a larger distance on their backside towards dominant others, when both border cases –those participants that did not believe the manipulation- and SDC are controlled for. No significant correlations between PPI-R total score and other proxemic scores are found, neither when border cases are included, or excluded. In table 2, results of regression analyses with PPI-R factor scores as predictors are reported. When PPI-R factor scores are used as predictors, and the analyses yield a significant correlation, factor III –Coldheartedness- is the only remaining predictor in the statistical model. Table 2 demonstrates that when SDC is controlled for, PPI-R factor III shows a significant positive correlation with behind distance and PS score. This indicates that individuals that score high on Coldheartedness show significantly larger distances towards others on their backsides, and show a significantly larger personal space overall. This relationship seems independent of whether the approaching individual is dominant or submissive, as behind distance and PS score appear significant in both the dominant and non-dominant condition.

Furthermore, table 2 demonstrates that factor III only positively correlates to right distance when it is assessed by a submissive interviewer. This finding differs per condition and therefore seems inconsistent, as it would indicate that individuals that score high on Coldheartedness show a larger distance towards others on their right sides, but not on their left sides. This relationship would also be dependent on the dominant/submissive nature of the individual that encroaches and/or whether border cases are included or excluded, and therefore might indicate a possible interaction effect. Additionally, table 2 demonstrates no significant correlation between PPI-R factors and front distance or left distance.

### **Effect of perceived dominance on Personal Space**

Hypothesis two tested whether PS is larger when it is assessed by a dominant male adult than when it is assessed by a non-dominant male adult. Table 3 shows the results of the performed paired sampled t-tests that compared proxemic scores. As a result of the confounding effect of the statistically differing confederate (SDC), a t-test was performed excluding SDC cases ( $N=51$ ). As table 3 demonstrates, front distance, left distance and PS score differ significantly, and behind distance and right distance do not differ significantly. This indicates that front distance, left distance and PS score tend to be larger –thus, distance between individuals is larger- when it is assessed by a dominant male adult, than when it is assessed by a non-dominant male adult. Subsequently, behind distance and right appear unaffected by dominance in male adults.



**Table 2:** Results of regression analyses, with PPI-R factors as predictors. PPI-R factor III –Coldheartedness– was the only remaining factor when the regression analysis yielded significant results. Numbers flagged with \* are significant at  $p < .05$ , numbers flagged with \*\* are significant at  $p < .01$ .

Dominant condition						
Step 3 Proxemic score	Model	Unstandardized		Standardized		
		B	SE B	$\beta$	T	p-value
Front	(Constant)	1.01	.80		1.37	.18
	PPI-III	.03	.02	.20	1.44	.16
Behind	(Constant)	-1.29	1.38		-.94	.35
	PPI-III	.07	.03	.31	2.25	.03*
Left	(Constant)	.93	.99		.93	.36
	PPI-III	.03	.02	.18	1.31	.20
Right	(Constant)	.71	1.04		.68	.50
	PPI-III	.04	.02	.21	1.48	.15
PS	(Constant)	-8.97	16.04		-.56	.58
	PPI-III	.75	.37	.28	2.05	.04*
Non-dominant condition						
Step 3 Proxemic score	Model	Unstandardized		Standardized		
		B	SE B	$\beta$	T	p-value
Front	(Constant)	.75	1.06		.71	.48
	PPI-III	.01	.01	.16	1.13	.27
Behind	(Constant)	-2.53	1.43		-1.78	.08
	PPI-III	.10	.03	.40	3.05	<.01*
Left	(Constant)	.31	1.06		.30	.77
	PPI-III	.04	.02	.23	1.67	.10
Right	(Constant)	-.35	1.03		-.35	.73
	PPI-III	.06	.02	.33	2.41	.02*
PS	(Constant)	-16.99	14.41		-1.18	.24
	PPI-III	.88	.33	.36	2.67	0.01*

**Table 3:** Results of the paired samples t-test. \* = statistically significant at  $p < .05$ , \*\* = statistically significant at  $p < .01$  and \*\*\* = statistically significant at  $p < .001$ .

Proxemic distance	Mean distance (SD) dominant condition	Mean distance (SD) non-dominant condition	T	p-value
Front	2.23 (.73)	1.93 (.71)	4.31	<.001***
Behind	1.78 (1.30)	1.77 (1.39)	.08	.94
Left	2.22 (.91)	2.07 (.98)	2.61	.01*
Right	2.25 (.95)	2.11 (.97)	2.01	.05
PS Score	23.61 (14.96)	21.13 (13.80)	2.63	.01*

Results on hypothesis one indicate that psychopathy does not seem directly correlated to proxemic scores, even when confounding cases are removed. However, Coldheartedness, a specific psychopathic trait, tends to be positively correlated with personal space and behind proxemic distance. This trait also appears uncorrelated with left distance. With regard to right proxemic distance, the analyses yielded varying results which was not in line with the original expectations of this study.

The results on hypothesis two appear consistent for at least total Personal Space score and front and behind distance. PS total score and front distance seem to correlate positively by perceived dominance in adult males whereas behind distance does not. This indicates that perceived dominance enlarges one's frontal distance and total personal space towards those perceived as dominant, whereas behind distance towards perceived dominant others does not. This is in line with the original hypothesis of this experiment, that PS would enlarge as a dominant individual would approach, compared to when a non-dominant individual would approach. This result might suggest that PS would only be influenced, when the approaching dominant individual is actually in the visual field. However, results on left and right proxemic distance appear to vary.

In the experiment, three different individuals fulfilled the role of dominant interviewer. A self-report manipulation check was performed to assess whether participants perceived these confederates significantly different on liking, skillfulness, willingness to be their employee and willingness to be their boss. No significant difference was measured. However, PS score did differ significantly among the actors, as one actor in particular differed significantly from the other two, where PS score was systematically larger. One plausible explanation could be that the actor's actual height (190cm) differed ten centimeters from the other two actors (180cm). This might have influenced the participant's perception of how close they would let a dominant individual approach them.

Since some proxemic scores differed significantly when assessed by the three different dominant confederates, this led to immediate methodological difficulties in testing the two hypotheses. The test on the hypothesis concerning psychopathic traits and PS also yielded varying results when the analysis was run excluding the

participants that had been tested by the statistically differing actor. When those participants were excluded, consistent results were found with regard to PS score and behind proxemic distance, as they were shown to be predicted by the level of Coldheartedness –a specific factor and trait of psychopathy. Front proxemic distance and left proxemic distance appear to be consistent as well, as they do not seem to show any correlation with psychopathy, or any of its' specific traits. Varying results are found with respect to right distance, as it varies depending on whether the participants were assessed by the dominant or submissive confederate. When they were assessed by the submissive confederate, a positive correlation indeed was found. The disadvantage of excluding the participants that had been tested by the statistically differing actor, is a drop in statistical power. When border cases and participants that were tested by the statistically differing actor were removed from the analyses, no significant correlation was observed between psychopathy and PS. This seems to be in line with results found by Eastwood (1985), and partly contradicts results found by Kinzel (1970), Newman & Pollack (1973) and Wormith (1984). However, a significant positive correlation was observed between psychopathy trait Coldheartedness and PS, as well as between the Coldheartedness trait and behind proxemic distance. The psychopathy trait Coldheartedness mildly predicts PS score in this sample. A possible explanation for this might be that a lack of empathic concern might lead to unawareness of the fact that maintaining a larger personal space might be unsociable towards others. Another explanation could be that a lack of empathic concern might cause individuals to maintain a larger buffer zone in which they could execute fight or flight behavior, regardless the intentions of approaching individuals. A third explanation might be that individuals who score high on Coldheartedness cared less about participating in this experiment, and therefore just responded faster on this particular task, but then one would expect random and non-significant results. This possibility would be even more plausible if abnormal responding was observed in the psychopathy self-report, but this was not the case. Apart from this, a lack of empathic concern might lead to a lack of interest in participating in this experiment. The present study tested a non-clinical population, while early effects were found within a clinical population (Kinzel, 1970, Newman & Pollack, 1973). The results of this study suggest that the relationship between personality and personal space can be observed more generally in both populations.

With respect to the results on the perceived dominance hypothesis, it was observed that front distance and PS score both were significantly larger when assessed by a dominant confederate, than when assessed by a non-dominant confederate. These results seem to vary across conditions, even when confounding border cases -cases that doubted the manipulation- and SDC cases -cases that were assessed by a statistically differing confederate- were controlled for. It must be noted that carryover effects between dominant and non-dominant conditions are not controlled, due to the methodological design. As theorized, these results support the notion that PS functions as an aggression regulator (Pfeifer, 1969). Dominant individuals could be seen as possibly more dangerous or aggressive. Inherently, PS is larger, and therefore a larger safety- or buffer zone is maintained. This effect was also found in earlier studies that tested forensic populations

(Kinzel 1970; Newman & Pollack, 1973; Wormith, 1984). It must also be noted that this effect could be enlarged due to the fact that only adult men were tested. Dominance and psychopathic personality traits manifests themselves more readily in men than women (Lilienfeld & Widows, 2005). Therefore, men could possibly react differently to dominance than women. With regard to the varying results found in left and right proxemic distance, a possible explanation could be the handedness of participants. As can be seen in the results, proxemic is significantly larger on the left hand side. This might supposedly indicate that participants could less easily defend themselves with their left hand. This might in turn be due to the fact that the left hand is supposedly the non-dominant hand in the majority of cases. However, this has not been measured in the experiment. Also, the latter is not supported by earlier studies (Kinzel, 1970; Newman & Pollack, 1973; Wormith, 1984; Eastwood, 1985). Taking into account the methodological limitations –relative low power and possible carryover effects, varying results due to possible handedness effects– of this study, future research could focus on possible height effects on PS. More importantly, when carryover and sex effects are controlled for, theoretical notions on both the effects of coldhearted personalities on PS, and the effects of dominance on PS, might become more plausible and evident. Also, varying results found in left and right PS distances then might be ruled out, as possible influencing factors such as handedness can be taken into account.

To conclude, this study finds varying results with respect to the correlation of psychopathic personality traits to personal space. Nevertheless, a particular psychopathic personality trait, Coldheartedness, does seem to mildly predict PS. This study's results therefore support possible personality effects on one's personal space in non-clinical populations. Subsequently, this study supports the notion that PS tends to enlarge when a dominant individual approaches, compared to when a non-dominant individual approaches.

### Acknowledgements

Experimental data used in this study is property of J. Lobbestael (PhD.), employee of Maastricht University, Maastricht, the Netherlands.

### REFERENCES

- Blackburn, R. (1975). An Empirical Classification of the Psychopathic Personality. *British Journal of Psychiatry*, 127, 456-460.
- Chauvenet, W. (1960). *A Manual of Spherical and Practical Astronomy V.II* (1st Ed (1863); Reprint of 1891 5th Ed ed.). Dover, N.Y.: Lippincott, Philadelphia.
- Cleckley, H. (1941). *The mask of sanity; an attempt to reinterpret the so-called psychopathic personality*. Oxford England: Mosby.
- Eastwood, L. (1985). Personality, Intelligence and Personal Space among Violent and Non-Violent Delinquents. *Personality and Individual Differences*, 6(6), 717-723.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *The Eysenck Personality Questionnaire*. London: Hodder & Stoughton.
- Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal Behavior and the Vertical Dimension

- of Social Relations: A Meta-Analysis. *Psychological Bulletin*, 131(6), 898-924.
- Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, 1(2), 111-119.
- Hare, R. D. (1991). *Manual for the Revised Psychopathy Checklist*. Toronto, Canada: Multi-Health Systems.
- Kinzel, A. F. (1970). Body-buffer zone in violent prisoners. *American Journal of Psychiatry*, 127, 59-64.
- Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, 66(3), 488-524.
- Lilienfeld, S. O., & Widows, M. R. (2005). Psychopathic personality inventory-revised: Professional manual. Lutz, FL: *Psychological Assessment Resources*.
- Morrison, D., & Gilbert, P. (2001). Social rank, shame and anger in primary and secondary psychopaths. *Journal of Forensic Psychiatry*, 12(2), 330-356.
- Newman, R. C., & Pollack, D. (1973). Proxemics in deviant adolescents. *Journal of Consulting and Clinical Psychology*, 40(1), 6-8.
- Pfeiffer, J. (1969). *The Emergence of Man*. New York: Harper & Row.
- Sandler, J. C. (2007). Computer equivalency of the Psychopathic Personality Inventory-revised in a nonincarcerated population. *Criminal Justice and Behavior*, 34(3), 399-410.
- Sommer, R. (1959). Studies in personal space. *Sociometry*, 22, 247-260.
- Uzieblo, K., Verschuere, B., Van den Bussche, E., & Crombez, G. (2010). The Validity of the Psychopathic Personality Inventory—Revised in a Community Sample. *Assessment*, 17(3), 334-346.
- Wormith, J. S. (1984). Personal space of incarcerated offenders. *Journal of Clinical Psychology*, 40(3), 815-827.

# The effects of mindfulness versus thought suppression instruction on the appraisal of emotional and neutral pictures

## ORIGINAL PAPER

The present study investigated the effects of brief mindfulness and thought suppression instructions on the appraisal of emotional (positive and negative) and neutral pictures. It was expected that mindfulness would promote positive emotional reactions toward picture stimuli, whereas thought suppression would promote negative emotional reactions. Sixty participants were randomly assigned to one of the three conditions (mindfulness, thought suppression, control). In each condition, participants rated emotional and neutral pictures on two 9-point scales, one pertaining to picture valence and the other to picture arousal. Contrary to what was expected, the present study revealed no effects of mindfulness on dealing with emotions provoked by emotional and neutral pictures. Thought suppression, on the other hand, was found to be a successful strategy in dealing with emotions provoked by negative pictures. It is suggested that thought suppression is a successful short-term emotion regulation strategy for dealing with negative emotions of low intensity.

**Keywords:** mindfulness; thought suppression; emotional pictures; neutral pictures

Marko Radivojevic; MSc in Health and Social Psychology  
Maastricht University, Maastricht, the Netherlands

---

[m.radivojevic@alumni.maastrichtuniversity.nl](mailto:m.radivojevic@alumni.maastrichtuniversity.nl)

## INTRODUCTION

Over the past couple of decades, the emergence of mindfulness has drawn a great deal of attention from behavioural researchers (Brown & Cordon, 2009; Kabat-Zinn, 1982). Definitions of mindfulness proposed by these researchers have often been diverse (Brown, Ryan, & Creswell, 2007; Follette, Palm, & Pearson, 2006), but regardless of this diversity, all of them use at least one of the concepts considered to be crucial for defining mindfulness: attention, awareness, present-centeredness, and non-judgmental stance toward mental content or acceptance (Baer, 2003; Brown et al., 2007; Germer, 2005).

Thought suppression has often been considered the opposite pole of mindfulness (Campbell-Sills, Barlow, Brown, & Hofmann, 2006a, 2006b; Hooper, Davies, Davies, & McHugh, 2011). Instead of being aware and accepting one's thoughts and feelings as they arise (i.e. being mindful), suppression involves conscious effort to ignore or deny the existence of an unwanted thought, which eventually results in an enhanced occurrence of that same thought (Hooper et al., 2011). Summing up literature on thought suppression, Wenzlaff & Wegner (2000) conclude that thought suppression is not merely an ineffective strategy of mental control; it is even counterproductive, fostering the state of mind one had initially hoped to avoid.

Emotional experience and its regulation are central to psychological well-being (Brown & Cordon, 2009). Successful regulation of emotions leads to improvements in overall well-being (Shapiro, Carlson, Astin, & Freedman, 2006), as well as social adjustment (Campbell-Sills et al., 2006b). Unsuccessful regulation of emotions, on the other hand, may lead to various mental health problems, which often have serious consequences for one's well-being (Follette et al., 2006). A number of studies aimed at investigating the effects of mindfulness and thought suppression on emotion regulation. With respect to mindfulness, studies point out that mindfulness helps to increase behavioural willingness and tolerance when dealing with negative material (Arch and Craske, 2006), attenuates emotional intensity when viewing highly emotional pictures in both experienced and beginner meditators (Taylor et al., 2011), and is found to be the best strategy when dealing with emotions provoked by emotional stimuli (Hooper et al., 2011). Studies of thought suppression, on the other hand, consistently claim that it is not only highly ineffective, but also counterproductive (Campbell-Sills et al., 2006b; Hooper et al., 2011; Hooper, Sandoz, Ashton, Clarke, & McHugh, 2012). Taken together, these studies suggest that the ability to accept emotions and to confront them with a non-judgmental stance apparently leads to a more successful emotion regulation and, consequently, improved overall well-being.

So far, the majority of studies conducted assessed the effects of aforementioned coping strategies on dealing with negative stimuli. It would be interesting to examine the effects of these strategies on positive and neutral stimuli as well, thus potentially providing some insight in the overall generalisability of the effects of these strategies across the whole spectrum of emotions. The present study has the purpose to add to the findings outlined above by comparing the effects of mindfulness versus thought



suppression instructions on the appraisal of positive, negative and neutral pictures. More precise, the aim of the present study is to explore the effects of mindfulness versus thought suppression instructions on the appraisal of positive, negative and neutral pictures. In line with the above mentioned studies, it was hypothesized that after hearing a mindfulness instruction, participants would evaluate the pictures as more pleasant and calming, whereas after a thought suppression instruction they would evaluate the pictures as less pleasant and calming. Since valence and arousal are considered to be fundamental dimensions of emotional experience (Bradley & Lang, 1994; Dolcos & Cabeza, 2002), hypotheses were made with respect to both picture valence and picture arousal. Specifically:

1. Mindfulness instruction will facilitate positive emotional reactions toward emotional (positive and negative) pictures, whereas it will have no significant effect on the appraisal of neutral pictures.
  - a) After a mindfulness instruction, positive pictures will be evaluated as more positive, negative pictures as less negative and neutral pictures will be evaluated as slightly more positive.
  - b) After a mindfulness instruction, both positive and negative pictures will be evaluated as less arousing, and neutral pictures will be evaluated as slightly less arousing.
2. Thought suppression instruction will facilitate negative emotional reactions to emotional (positive and negative) pictures, whereas it will have no significant effect on the appraisal of neutral pictures.
  - a) After a thought suppression instruction, positive pictures will be evaluated as less positive, negative pictures as more negative, and neutral pictures will be evaluated as slightly more negative.
  - b) After a thought suppression instruction, both positive and negative pictures will be evaluated as more arousing, and neutral pictures will be evaluated as slightly more arousing.

## METHODOLOGY

### Participants

Sixty undergraduate and graduate Maastricht University students enrolled in various study programs participated in this experiment in return for either course credit or money vouchers. They were randomly assigned to one of the three conditions: mindfulness ( $n=20$ ), thought suppression ( $n=20$ ) or control ( $n=20$ ).

### Design

The present study is part of a more extensive study that also included investigation of the effects of emotional and neutral pictures on memory (recall/recognition task). Overall, the study had four main dependent variables, two of which were the focus of this research - mean valence ratings and mean arousal ratings. Independent variables were the three different conditions used: mindfulness instruction, thought suppression instruction and no instruction (control). A 3 (condition: mindfulness,

thought suppression, no instruction) by 3 (valence ratings: positive, negative, neutral pictures) by 3 (arousal ratings: positive, negative and neutral pictures) between-subject mixed factorial design was used.

### **Stimuli**

In total, 120 pictures taken from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) were used as stimulus material. Each picture category used in this research (positive, negative and neutral) included 40 pictures. Sixty of those pictures were used in the rating task, 20 per category. For each category, pictures were selected based on their mean valence and arousal ratings (i.e. norm data). According to Dolcos and Cabeza (2002), valence refers to a continuum ranging from pleasant to unpleasant, with neutral as an intermediate value, whereas arousal pertains to a continuum ranging from calm to excitement. Positive pictures had a mean valence rating of above 7 and a mean arousal rating of above 5 (both on a 9-point scale). Negative pictures had mean ratings of below 3 for valence and above 5 for arousal. Mean valence ratings for neutral pictures ranged between 3.1 to 7 and mean arousal ratings for those pictures were equal to or below 5.

### **Procedure and materials**

In order not to reveal the true purpose of the study to the participants, the experiment was presented as a simple task of evaluating emotional pictures on dimensions of valence and arousal. After arriving at the lab, participants were provided with an informed consent form, which they had to sign in order to start the testing.

*Mood* - In an attempt to see whether the effects of the instructions could be separated from the effects of mood, participants' mood was assessed using a 16-item Brief Mood Inspection Scale (BMIS; Mayer & Gaschke, 1988). The scale consists of 5-point items ranging from 1 (*very slightly or not at all*) to 5 (*extremely/ totally*). Participants had to indicate to what extent each statement from the scale applied to them. They were assessed two times – once before the onset of the actual experimental procedure (i.e. baseline measurement) and once after they had heard the emotion regulation instruction (i.e. mindfulness or thought suppression). The majority of other similar studies (e.g. Alberts & Thewissen, 2011) report no influence of mood on dependent variables.

*Mindfulness* - Next, participants' level of dispositional mindfulness (trait mindfulness) was measured using the Mindful Attention Awareness Scale (MAAS; Brown & Ryan, 2003). MAAS consists of 15 6-point items (ranging from 1 – *almost always* to 6 – *almost never*) measuring attention to, and awareness of what occurs in the present moment. Internal consistency ( $\alpha$ ) of the scale was found to be .82 in the student sample and .87 in the adult sample. MAAS was used to assess whether there were any pre-experimental differences in mindfulness among different groups of participants. In addition, participants were asked to indicate their level of meditation experience using one 3-point item ranging from 1 (*no meditation experience*) to 3 (*considerable meditation experience*).

*Instructions* - After participants had filled in the BMIS and MAAS, they received mindfulness or thought suppression instructions, depending on the experimental

group they were in. The audiotaped instructions were equal in length (5 minutes) and approximate number of words used. The control group received no specific instructions with respect to coping with the upcoming stimuli.

*Rating task* - After listening to the instructions, participants began with a rating task. The rating task consisted of 20 positive, 20 negative and 20 neutral pictures. Participants were instructed to rate each picture on two 9-point scales by pressing the appropriate number on a keyboard (1 to 9), indicating to what extent the picture is pleasant/unpleasant (valence rating) and arousing/calming (arousal rating) to them personally. This 9-point scale, taken from Bradley and Lang (1994), represents a picture-oriented scale called Self-Assessment Manikin (SAM). SAM ranges from a happy, smiling figure to an unhappy and frowning figure when valence is assessed, and from excited, open-eyed figure to a calm, sleepy figure when representing arousal. Every picture was presented on the screen for 3 seconds and participants rated it on valence and arousal immediately after seeing it. Participants from the experimental groups were instructed to apply their respective instructions when dealing with the stimuli. Control participants were told to rate the pictures based on their first impression.

*Filler task* - A filler task took place after the rating task and before the next, recall/recognition task. The computer game Tetris Unlimited (Martinez, 2003) was chosen as a filler task. This game draws on the ability of mental rotation and is unlikely to affect the memory of previously viewed pictures (Alberts & Thewissen, 2011). Participants played Tetris Unlimited for 20 minutes.

*Recall/recognition task* - Next, participants engaged in a recall/recognition task. Since recall/recognition task is not the focus of the current study, it will not be further discussed here.

*Manipulation check* - After the testing phase, participants from mindfulness and thought suppression groups received a short instruction application questionnaire. This questionnaire consisted of 2 questions – one asked participants if they had been successful in applying the instructions (yes/no question), whereas the other was a 10-point item asking participants to rate their success in applying the instruction. The control group, on the other hand, received a questionnaire asking them to indicate whether they had used some particular strategy when confronted with the pictures and if they did, which strategy it was.

At the end of the experiment, participants were thanked for their participation and told that they would receive a debriefing form after all the participants had been tested.

## **Statistical analyses**

### *Mood*

Repeated measures ANOVA with the score on BMIS as within subjects factor (before and after the instruction) and condition (mindfulness, thought suppression and control) as between subjects factor was used to assess differences in mood between groups on the two measurements

### *Mindfulness*

One-way ANOVA with condition as independent variable and scores on the MAAS as dependent variable was used to assess levels of dispositional mindfulness in different groups of participants. Participants' meditation experience was also assessed using one-way ANOVA with condition as independent variable and scores on the 3-point meditation experience item as dependent variable.

### *Instructions and strategy use*

Scores on the instruction and strategy use questionnaire were computed and then compared using independent samples t-test.

### *Valence and arousal*

Repeated measures ANOVA with condition as between subjects factor and mean valence and arousal ratings respectively as within subjects factors (both with three levels: positive, negative and neutral pictures) was conducted to examine the effect of a specific instruction (or lack thereof) on participants' picture evaluation. Additional multivariate ANOVAs were conducted to further investigate interaction effects between the two factors.

## RESULTS

### **Mood**

Repeated measures ANOVA revealed a significant main effect of mood,  $F(1, 38) = 83.01, p = .00$ . After hearing the instructions, participants' scores were significantly lower ( $M = 3.56, SD = .38$ ) than before ( $M = 3.90, SD = .46$ ), irrespective of the group they were in. However, both main effect of condition ( $F(1, 38) = .54, p = .47$ ) and the interaction effect ( $F(1, 38) = 2.48, p = .12$ ) were found to be non-significant. These results indicate that the reported values of valence and arousal ratings may be affected by changes in participants' mood due to the instructions, and not by the instructions per se.

### **Mindfulness**

One-way ANOVA did not yield a significant effect of mindfulness,  $F(2, 57) = 1.21, p = .31$ . This result indicates that participants in mindfulness ( $M = 56.80, SD = 8.21$ ), thought suppression ( $M = 58.95, SD = 10.07$ ) and control ( $M = 54.55, SD = 8.43$ ) conditions did not differ with respect to their dispositional levels of mindfulness. Analysis of meditation experience also revealed a non-significant effect,  $F(2, 57) = 2.81, p = .07$ . Participants in mindfulness ( $M = 1.10, SD = .31$ ), thought suppression ( $M = 1.45, SD = .61$ ) and control condition ( $M = 1.25, SD = .44$ ) did not differ in their levels of prior meditation experience.

### Instructions and strategy use

Overall, 85% of participants from the experimental groups reported success in applying the instructions. In the mindfulness group, 90% of participants reported being successful, whereas in the thought suppression group 80% of participants successfully applied the instruction. The independent samples t-test results revealed that both mindfulness ( $M = .90$ ,  $SD = .31$ ) and thought suppression ( $M = .80$ ,  $SD = .41$ ) groups were equally successful in applying their respective instructions,  $t(38) = .87$ ,  $p = .39$ . Moreover, both groups (mindfulness,  $M = 7.10$ ,  $SD = 1.25$ ; thought suppression,  $M = 6.75$ ,  $SD = 2.05$ ) rated the success in applying these instructions equally,  $t(38) = .65$ ,  $p = .52$ .

Although control group participants were theoretically allowed to use multiple emotion regulation strategies while viewing the pictures, the majority did not report using any strategy (60%). Of the participants who reported using some strategy, 40% tried not to think about the content of the pictures and/or to view the pictures without the emotional charge. Thirty percent tried to think about something other than the pictures. Finally, 20% of participants tried to look away from the pictures and/or used some other strategy, not listed in the questionnaire, for dealing with the emotions that the pictures provoked.

Participants' mean mood, mindfulness and instruction questionnaire scores are shown in Table 1.

**Table 1.** Means (Standard Deviations) for Mood, Mindfulness and Instruction Application Scales. <sup>a</sup> Brief mood inspection scale administered before the instruction. <sup>b</sup> Brief mood inspection scale administered after the instruction. <sup>c</sup> Mindful attention awareness scale. <sup>d</sup> Meditation experience. <sup>e</sup> Success in applying the instructions. <sup>f</sup> Success in applying the instruction rating.

Condition	Mood		Mindfulness		Instructions	
	BMIS 1 <sup>a</sup>	BMIS 2 <sup>b</sup>	MAAS <sup>c</sup>	ME <sup>d</sup>	SAI <sup>e</sup>	SAIR <sup>f</sup>
Mindfulness	3.83 (.48)	3.54 (.37)	56.80 (8.21)	1.10 (.31)	.90 (.31)	7.10 (1.25)
Thought suppression	3.98 (.43)	3.58 (.39)	58.95 (10.07)	1.45 (.61)	.80 (.41)	6.75 (2.05)
Control	--	--	54.55 (8.43)	1.25 (.44)	--	--
<b>Total</b>	3.90 (.46)	3.56 (.38)	56.77 (8.98)	1.27 (.48)	.85 (.36)	6.93 (1.69)

Note. Control participants were not included in the analyses of mood and instruction application data.

### Valence and Arousal

Since in both repeated measures ANOVAs Mauchly's test indicated that the assumption of sphericity had been violated (for ANOVA on valence ratings,  $\chi^2(2) = 32.22$ ,  $p = .00$ ; for ANOVA on arousal ratings,  $\chi^2(2) = 26.08$ ,  $p = .00$ ), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (for ANOVA on valence ratings,  $\epsilon = .70$ ; for ANOVA on arousal ratings,  $\epsilon = .73$ ). Mean valence and arousal ratings for mindfulness, thought suppression and control conditions for each picture category are shown in Table 2.

**Table 2.** Means (Standard Deviations) of Valence and Arousal Ratings for Mindfulness, Thought Suppression and Control Condition

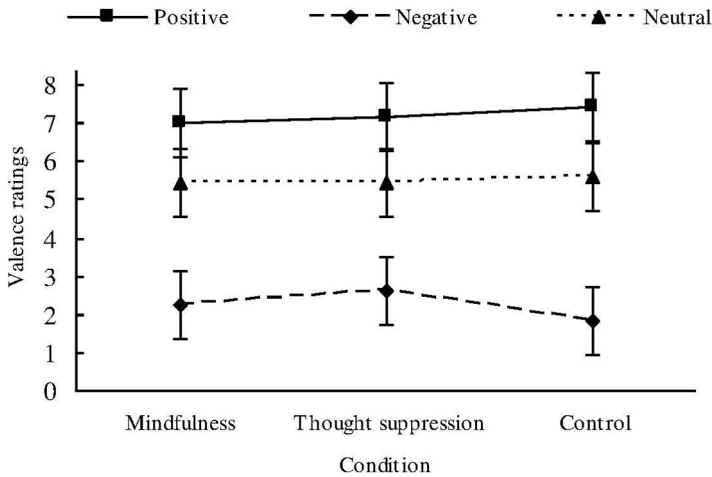
	<i>Condition</i>	<i>Valence ratings</i>	<i>Arousal ratings</i>
Positive pictures	Mindfulness	7.02 (.66)	5.09 (1.53)
	Thought suppression	7.17 (.86)	5.06 (1.72)
	Control	7.41 (.83)	5.62 (2.07)
	Total	7.20 (.79)	5.25 (1.78)
Negative pictures	Mindfulness	2.24 (.68)	5.77 (1.68)
	Thought suppression	2.61 (.96)	4.27 (1.82)
	Control	1.82 (.62)	5.59 (1.46)
	Total	2.22 (.82)	5.21 (1.77)
Neutral pictures	Mindfulness	5.43 (.52)	3.07 (1.21)
	Thought suppression	5.44 (.30)	3.01 (1.23)
	Control	5.62 (.65)	3.26 (1.43)
	Total	5.49 (.51)	3.11 (1.28)

### *Valence*

Repeated measures ANOVA with condition as between subjects factor and mean valence ratings as within subjects factor revealed a significant interaction between the two factors,  $F(2.78, 79.31) = 3.66, p = .02$ . A follow-up multivariate ANOVA was conducted, revealing significant differences between conditions, but only with respect to negative picture ratings,  $F(2, 57) = 5.36, p = .01$ . No significant differences between the three conditions were found with respect to positive ( $F(2, 57) = 1.23, p = .30$ ) and neutral ( $F(2, 57) = .88, p = .42$ ) picture ratings. Further post hoc tests indicated that participants in the thought suppression condition ( $M = 2.61, SD = .96$ ) rated negative pictures as more pleasant compared to participants in the control condition ( $M = 1.82, SD = .62$ ). There were no significant differences in negative picture ratings between mindfulness ( $M = 2.24, SD = .68$ ) and thought suppression condition, nor between mindfulness and control condition. Figure 1 depicts valence ratings for the three conditions and shows the difference in valence ratings between the thought suppression and control condition within the negative pictures category.

In addition to conducting the analyses on original data, both repeated measures and multivariate ANOVA were conducted on the data set containing no outliers, i.e. participants who had valence ratings that were more than 2.5 standard deviations higher or lower than the mean score. There were two such participants in the present study. It has been argued that outliers may lead to distortions of parameter and statistic estimates (Zimmerman, 1994). In the present research, however, no changes in the pattern of results occurred when outliers were excluded from the analyses. Repeated measures ANOVA revealed a significant interaction effect,  $F(2.84, 78.03) = 4.59, p = .01$ . Multivariate analyses of simple effects repeated the same pattern of results obtained in the original analyses – significant differences between

conditions were found only with respect to negative picture ratings,  $F(2, 55) = 6.44$ ,  $p = .00$ . There were no differences between conditions neither in positive ( $F(2, 55) = 1.64$ ,  $p = .20$ ) nor in neutral ( $F(2, 55) = .42$ ,  $p = .66$ ) picture category. Further post hoc tests indicated that participants in the thought suppression condition ( $M = 2.61$ ,  $SD = .96$ ) evaluated negative pictures as more pleasant compared to participants in the control condition ( $M = 1.74$ ,  $SD = .53$ ). There were no significant differences in negative picture ratings between mindfulness ( $M = 2.24$ ,  $SD = .68$ ) and thought suppression condition, nor between mindfulness and control condition.



**Figure 1:** Self-reported valence ratings of positive, negative and neutral pictures for participants from mindfulness, thought suppression and control groups.

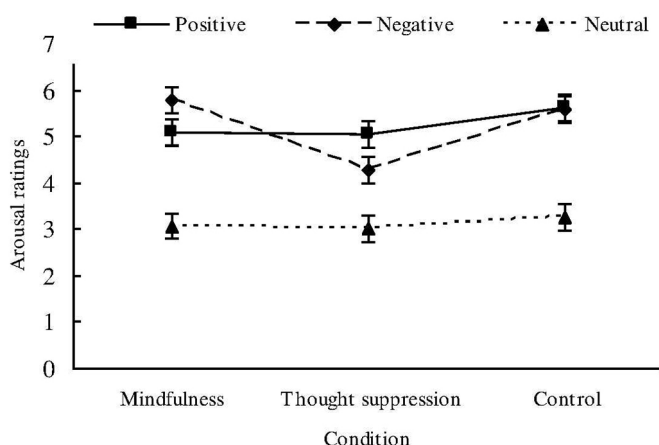
### Arousal

A similar repeated measures ANOVA procedure conducted for arousal ratings did not reveal a significant interaction effect,  $F(2.92, 83.07) = 2.43$ ,  $p = .07$ . However, since the interaction was only slightly insignificant, multivariate ANOVA was conducted to further investigate interaction effects. The analysis conducted after the exclusion of outliers from the initial analysis served as justification for the use of multivariate ANOVA on the initial results. In the analyses of outliers, three participants were found to have arousal ratings that were more than 2.5 standard deviations higher or lower than the mean score. The repeated measure ANOVA without these outliers revealed a significant interaction effect,  $F(3.06, 82.59) = 4.42$ ,  $p = .01$ . A follow up multivariate ANOVA revealed significant differences between conditions, but only with respect to negative picture ratings,  $F(2, 54) = 8.44$ ,  $p = .00$ . No significant differences between the three conditions were found with respect to positive ( $F(2, 54) = .98$ ,  $p = .38$ ) and neutral ( $F(2, 54) = .23$ ,  $p = .79$ ) picture ratings. Further post hoc tests indicated that participants in the thought suppression condition ( $M = 4.27$ ,  $SD = 1.82$ ) rated negative pictures as less arousing compared to participants in both mindfulness ( $M = 5.97$ ,  $SD = 1.45$ ) and control condition ( $M = 5.94$ ,  $SD$



= 1.03). There were no significant differences in negative picture ratings between mindfulness and control condition.

When multivariate ANOVA was conducted on the initial results, the same pattern was revealed. Significant differences between conditions were found only with respect to negative picture ratings,  $F(2, 57) = 4.89, p = .01$ . No significant differences between the three conditions were found with respect to positive ( $F(2, 57) = .61, p = .55$ ) and neutral ( $F(2, 57) = .20, p = .82$ ) picture ratings. Post hoc tests further indicated that participants in the thought suppression condition ( $M = 4.27, SD = 1.82$ ) rated negative pictures as less arousing compared to participants in both mindfulness ( $M = 5.77, SD = 1.68$ ) and control condition ( $M = 5.59, SD = 1.46$ ). There were no significant differences in negative picture ratings between mindfulness and control condition. Figure 2 depicts arousal ratings for the three conditions and shows the difference in arousal ratings between the thought suppression condition and mindfulness and control conditions within the negative pictures category.



**Figure 2:** Self-reported arousal ratings of positive, negative and neutral pictures for participants from mindfulness, thought suppression and control groups.

## DISCUSSION

The current study examined the effects of mindfulness and thought suppression instructions on the appraisal of emotional (positive and negative) and neutral pictures. It was hypothesized that participants in the mindfulness condition would evaluate both emotional and neutral pictures as more positive compared to participants in the control group, whereas participants in the thought suppression condition would evaluate the pictures as more negative. The results, however, provided weak support for these assumptions. It was concluded that thought suppression is a successful strategy for dealing with emotions provoked by negative-valenced pictures. On the other hand, mindful attention to either positive, negative

or neutral material did not have any effect on dealing with emotions provoked by such material.

Further analyses were conducted in order to control for participants' mood, trait mindfulness and success in applying the instructions. Analysis of mood revealed that participants' mood changed after hearing their respective instructions. It is therefore possible that participants' valence and arousal ratings were not directly influenced by the instructions, but instead by changes in their mood after hearing the instructions. Analysis of trait mindfulness revealed no differences between participants in the three different conditions with respect to trait mindfulness and meditation experience. Finally, 85% of participants from the experimental groups reported success in applying their instructions. Furthermore, the majority (60%) of the control group participants reported not using any particular strategy when viewing the pictures, whereas of the ones that did use some strategy, majority (40%) tried not to think about the content of the pictures and/or to view the pictures without the emotional charge.

From a standpoint of the prevalence of literature on the topic, this study's findings regarding the effects of mindfulness and thought suppression are unexpected. Both quantitative and qualitative studies on mindfulness have repeatedly shown that mindfulness has a positive impact on emotions, especially with respect to negative emotions (Brown et al., 2007; Holzel et al., 2011; Nickerson & Hinton, 2011). However, Lykins and Baer (2009) suggested that in order for these beneficial effects of mindfulness to take place, one has to be an experienced mindfulness practitioner. Therefore, despite the fact that brief mindfulness instructions were shown to yield desired effects (e.g., Hooper et al., 2011), it is still conceivable that the lack of time to really grasp the instruction to stay with the emotion, as well as relative meditation inexperience might have contributed to potential difficulties in understanding the complex concept of mindfulness.

Another explanation for not finding the desired mindfulness effects concerns the picture stimuli used in the experiment. It is possible that the content of the pictures was too distant and not meaningful enough for participants to really experience the emotion. In other words, it might be that, for the actual effects of mindfulness to take place, the material has to be emotionally salient and relevant to the self, at least if one is not an experienced mindfulness practitioner.

The most striking finding of the present study concerns thought suppression. When instructed to suppress their emotions elicited by negative pictures, participants evaluated the pictures as more pleasant and less arousing. This finding is in opposition to the majority of evidence presented in classical studies on the subject (e.g. Wegner, Schneider, Carter III, & White 1987). However, findings have emerged in which this one-sided interpretation of the effects of suppression is called into question. Not only do these studies show that suppression leads to a decrease of expressive behavior (Goldin, McRae, Ramel, & Gross, 2008; Gross & Levenson, 1997; Jackson, Malmstadt, Larson, & Davidson, 2000), but they also indicate that suppression is a successful strategy for reducing distress and other forms of negative subjective experience (Goldin et al., 2008; Pilecki & Mckay, 2012). To account for this finding, Pilecki and Mckay (2012) suggested that suppression could actually be a successful *short-term* strategy for dealing with emotional stimuli

that are presented in a limited period of time. Abramowitz, Tolin and Street (2001) added that suppression is efficient short-term because of the absence of immediate surge of suppressed thoughts into one's mind, i.e. immediate enhancement effect. However, after a certain thought-suppression period, people begin to experience resurgence of those thoughts, i.e. a rebound effect, which consequently leads to a greater feeling of discomfort and the emergence of negative emotions. It is therefore possible that the amount of time participants spent suppressing emotions in the present research was not sufficient for a rebound effect to take place. As a result, participants possibly made a good use of the thought suppression strategy and were able to successfully manage emotions elicited by negative pictures. This implies that suppression might be a successful initial strategy for handling traumatic events. However, after a certain adjustment period, suppression should be replaced with some regulation strategy that potentially has more favorable long-term effects.

Another potential explanation of this study's findings regarding the benefits of thought suppression is that suppression may be effective for negative thoughts of low intensity. Pilecki and Mckay (2012) argue that low-intensity thoughts are less relevant to the self and are, hence, more easily suppressed. It might be that the stimuli used in the present study were not close enough to participants' experience in order to be perceived as highly intense, which aided the effort to suppress the emotions provoked by such stimuli.

One unexpected finding of the present study was that participants' mood was different after hearing the instructions. Particularly, they felt worse after they had followed their respective instructions. Although there are no studies dealing with the impact of mood state on mindfulness, there is some literature investigating the relation between mood and thought suppression (Purdon and Clark, 2001; Wyland and Forgas, 2007). Research has shown that people are better able to suppress unwanted thoughts when in a negative mood state (Wyland and Forgas, 2007). Wyland and Forgas (2007) stress that their suppression task was limited to a short period of time and targeted only neutral thoughts and not individually salient thoughts. Taken together, these findings are consistent with the present study's conclusions. The result of the present study, indicating that participants' mood was worse than it was before they heard the instructions, supports the speculations on the effectiveness of thought suppression as a short-term emotion regulation strategy. In spite of these considerations, the impact of mood on thought suppression and other emotion regulation strategies remains unclear and should be further explored.

One limitation of the study pertains to the potential impersonal nature of the stimuli used in the study. It is possible that participants tried to apply mindfulness and thought suppression instructions with a varying degree of effort due to the personal irrelevance of the emotions provoked by the stimuli, thus distorting the results. Another limitation concerns the number of stimuli used in the study. It has been suggested that thought suppression may only be effective in a short period of time. In the present study, 60 stimuli were used and participants did not spend longer than approximately 10 minutes trying to suppress their emotions. It is possible that in everyday life people spend a lot more time suppressing different emotions, which eventually leads to the occurrence of a rebound effect and failure

to self-regulate. Another methodological limitation pertains to the exclusive use of self-report measures for evaluation of valence, arousal and participants' success in applying the instructions. Campbell-Sills, Barlow, Brown & Hoffman (2006a) argue that self-reports are only indicative of attitudes and behaviors that are conscious and add that much emotion regulation takes place outside of conscious awareness. Future studies should therefore include some implicit measures of the degree of emotional intensity provoked by the stimuli and also find a way to measure participants' adherence to the instructions more reliably and objectively.

To conclude, this study adds to the literature that casts doubt over what was thought to be almost unquestionable – that mindfulness is highly effective and thought suppression highly ineffective emotion regulation strategy, and therefore calls for researchers to specify the conditions under which both of these strategies are most effective.

## REFERENCES

- Abramowitz, J., S., Tolin, D., F. & Street, G., P. (2001). Paradoxical effects of thought suppression: A meta-analysis of controlled studies. *Clinical Psychology Review*, 21, 683-703.
- Alberts H. J. E. M., & Thewissen, R. (2011). The effect of a brief mindfulness intervention on memory for positively and negatively valenced stimuli. *Mindfulness*, 2, 73-77.
- Arch, J. J., & Craske, M. G. (2006). Mechanisms of mindfulness: Emotion regulation following a focused breathing induction. *Behaviour Research and Therapy*, 44, 1849-1858.
- Baer, R.A. (2003). Mindfulness training as a clinical intervention: A conceptual and empirical review. *Clinical Psychology: Science and Practice*, 10, 125-143.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49-59.
- Brown, K. W., & Cordon, S. (2009). Toward a phenomenology of mindfulness: Subjective experience and emotional correlates. *Clinical Handbook of Mindfulness*, 2, 59-81.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84, 822-848.
- Brown, K.W., Ryan, R.M., & Creswell, J.D. (2007). Mindfulness: Theoretical foundations and evidence for its salutary effects. *Psychological Inquiry*, 18, 211-237.
- Campbell-Sills, L., Barlow, D. H., Brown, T. A., & Hofmann, S. G. (2006a). Acceptability and suppression of negative emotion in anxiety and mood disorders. *Emotion*, 6, 587-595.
- Campbell-Sills, L., Barlow, D. H., Brown, T. A., & Hofmann, S. G. (2006b). Effects of suppression and acceptance on emotional responses of individuals with anxiety and mood disorders. *Behavior Research and Therapy*, 40, 1251-1262.
- Dolcos, F., & Cabeza, R. (2002). Event-related potentials of emotional memory: Encoding pleasant, unpleasant, and neutral pictures. *Cognitive, Affective and Behavioral Neuroscience*, 2, 252-263.
- Follette, V., Palm, K. M., & Pearson, A. N. (2006). Mindfulness and trauma: Implications for treatment. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 24, 45- 61.
- Germer, C. K. (2005). Mindfulness: What is it? What does it Matter. In C. K. Germer, Siegel, R. D., & Fulton, P. R., (Eds.), *Mindfulness and psychotherapy* (pp. 3-27). New York: The Guilford Press.
- Goldin, P. R., McRae, K., Ramel, W. & Gross, J. J. (2008). The neural bases of emotion regulation: Reappraisal and suppression of negative emotion. *Biological Psychiatry*, 63,

577-586.

- Gross, J. J. & Levenson, R. W. (1997). Hiding feelings: The acute effects of inhibiting negative and positive emotion. *Journal of Abnormal Psychology*, 106, 95-103.
- Holzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R. & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science*, 6, 537-559.
- Hooper, N., Davies, N., Davies, L., & McHugh, L. (2011). Comparing thought suppression and mindfulness as coping techniques for spider fear. *Consciousness and Cognition*, 20, 1824-1830.
- Hooper, N., Sandoz, E. K., Ashton, J., Clarke, A., & McHugh, L. (2012). Comparing thought suppression and acceptance as coping techniques for food cravings. *Eating Behaviors*, 13, 62-64.
- Jackson, D. C., Malmstadt, J. R., Larson, C. L. & Davidson, R. J. (2000). Suppression and enhancement of emotional responses to unpleasant pictures. *Psychophysiology*, 37, 515-522.
- Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, 4, 33-47.
- Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A8*. University of Florida, Gainesville, FL.
- Lykins, E. L. B., Baer, R. A. (2009). Psychological functioning in a sample of long term practitioners of mindfulness meditation. *Journal of Cognitive Psychotherapy: An International Quarterly*, 23, 226-241.
- Martinez, O. G. (2003). Tetris Unlimited (v0.5.0) [Computer software]. Cambridge, MA.
- Mayer, J. D., & Gaschke, Y. N. (1988). The experience and metaexperience of mood. *Journal of Personality and Social Psychology*, 55, 102-111.
- Nickerson, A. & Hinton, D. E. (2011). Anger regulation in traumatized cambodian refugees: The perspectives of buddhist monks. *Culture, Medicine and Psychiatry*, 35, 396-416.
- Pilecki, B. C. & McKay, D. (2012). An experimental investigation of cognitive defusion. *The Psychological Record*, 62, 19-40.
- Purdon, C. & Clark, D. A. (2001). Suppression of obsession-like thoughts in nonclinical individuals: Impact on thought frequency, appraisal and mood state. *Behaviour Research and Therapy*, 39, 1163-1181.
- Shapiro, S. L., Carlson, L. E., Astin, J. A., & Freedman, B. (2006). Mechanisms of mindfulness. *Journal of Clinical Psychology*, 62, 373-386.
- Taylor, V. A., Grant, J., Daneault, V., Scavone, G., Breton, E., Roffe-Vidal, S., ... Beauregard, M. (2011). Impact of mindfulness on the neural responses to emotional pictures in experienced and beginner meditators. *Neuroimage*, 57, 1524-1533.
- Wegner, D. M., Schneider, D. J., Carter, S. R. & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, 53, 5-13.
- Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, 51, 59-91.
- Wyland, C. L. & Forgas, J. P. (2007). On bad mood and white bears: The effects of mood state on ability to suppress unwanted thoughts. *Cognition and Emotion*, 21, 1513-1524.
- Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. *The Journal of General Psychology*, 121, 391-401.

# Functional and anatomical connectivity in Complex Regional Pain Syndrome: A Multi-modal magnetic resonance imaging study

## ORIGINAL PAPER

Pain processing relies on a distributed network of cortical regions. It is therefore proposed that cortical connectivity analyses may be useful for examining the abnormal cortical network interactions that modulate pain perception, and other symptoms, in chronic pain conditions. The present study used functional magnetic resonance imaging and diffusion weighted magnetic resonance imaging to investigate differences in functional and anatomical connectivity between healthy controls and patients with complex regional pain syndrome (type 1). It was hypothesised that this chronic pain condition would associate with changes in functional connectivity within resting-state networks, as well as changes in fractional anisotropy in white matter tracts. Results suggested that there is increased functional connectivity between regions of the sensorimotor resting-state network, and that there is decreased fractional anisotropy in the portion of the superior longitudinal fasciculus connecting these regions. These regions correspond to primary and secondary somatosensory cortices, which have been implicated in the sensory-discriminative aspect of pain processing. This supports the hypothesis that there are both functional and anatomical changes within pain processing networks, in patients with complex regional pain syndrome.

**Keywords:** chronic pain; connectivity; resting-state; fractional anisotropy

Emma E. Biggs; Master student Cognitive Neuroscience  
Maastricht University, Maastricht, the Netherlands

---

[e.biggs@student.maastrichtuniversity.nl](mailto:e.biggs@student.maastrichtuniversity.nl)



## INTRODUCTION

Complex regional pain syndrome (CRPS) is a chronic pain disorder with a wide range of symptoms—that typically persist for over a year. There are two different types of CRPS: type I (CRPS-I; formerly known as reflex sympathetic dystrophy) which is preceded by mild tissue injury but no nerve damage, and type II (CRPS-II; formerly known as causalgia) which follows a nerve lesion (Patterson, Li, Smith, Smith, & Koman, 2011). In the Netherlands, incidence rates for CRPS are estimated to be 26.3 cases per 100,000 person-years (De Mos et al., 2007). The only identifiable risk factor appears to be limb immobilisation, and studies pre-CRPS show no differences in psychological state between those who do and do not develop CRPS (Marinus et al., 2011).

The main symptom of CRPS is spontaneous, diffuse pain, which covers an area larger than the area that suffered the initial trauma (De Boer et al., 2011). This pain is often described as a tearing or burning sensation (Patterson et al., 2011) that is not relieved by narcotics (Harke, Gredenort, Ladleif, Rahman, & Harke, 2001), and spreads proximally as the disorder persists (Marinus et al., 2011). In addition to these symptoms, there are a wide range of physiological signs associated with CRPS; sensory signs include increased sensitivity to tactile stimulation (hyperesthesia) and to painful stimuli (hyperalgesia), as well as the perception of non-painful stimuli as extremely painful (allodynia). The affected limb may also display vasomotor (colour and temperature) changes, sudomotor signs (abnormal sweating), motor signs (decreased strength and slowness of movement) and oedema (swelling) (Maihöfner, Handwerker, & Birklein, 2006).

Changes in cognition and emotion have also been documented: studies have shown that patients mistakenly perceive sensations on their affected limb when tactile stimuli are applied to their unaffected limb (referred sensations; Maihöfner, Neundörfer, Birklein, & Handwerker, 2006). In addition, some patients report feeling that their affected limb does not belong to them (cognitive neglect; Swart, Stins, & Beek, 2009). In terms of emotional changes, patients often develop anxiety and a fear of the pain they experience as well as low mood or depression (Patterson et al., 2011).

The interaction between the peripheral nervous system, biochemical processes and implicated cortical networks produce these physiological, cognitive and emotional symptoms. For example, the trauma initiates an aberrant immune response (mediated by increased levels of cytokines) which leads to neurogenic inflammation, and this sensitises central and peripheral nervous systems (Alexander, Peterlin, Perreault, Grothusen, & Schwartzman, 2011; Kasper et al., 2005). These processes explain a number of the physiological symptoms seen in CRPS-I patients, such as: oedema, temperature changes, sudomotor changes, hyperesthesia, and hyperalgesia. Maladaptive neuroplasticity in white and grey matter may cause the other symptoms, such as allodynia, referred sensations, and anxiety.

In normal pain processing, the motor and somatosensory cortices are functionally connected with the posterior insular cortex—forming a sensory-discriminative network of pain processing. The affective-motivational network of pain processing includes the anterior insular cortex, the prefrontal cortex and



anterior cingulate cortex (Peltz et al., 2011). It should be noted that these regions are not specific to pain processing, but are thought to modulate the perception of pain as part of a wider network of areas that also show consistent activation during painful stimulation (Iannetti & Mouraux, 2010).

Research has shown that in CRPS-I patients, the density of the anterior insular cortex is decreased (Baliki, Schnitzer, Bauer, & Apkarian, 2011a) and its atrophy is negatively correlated to pain duration (Geha et al., 2009). The study by Geha et al. (2009) demonstrated that this applies to the right anterior insular cortex (which represents autonomic and visceral responses), irrespective of the sidedness of the affected limb. Furthermore, the right ventromedial pre-frontal cortex, which is crucial to emotional decision-making, has decreased grey matter density in patients with CRPS-I, and its decrease correlates negatively with pain duration and severity (Geha et al., 2009). These regions (the right anterior insular cortex and ventromedial pre-frontal cortex) are more connected in patients with CRPS-I, than in healthy controls (Geha et al., 2009), supporting the theory that there is altered affective-motivational pain processing in CRPS-I patients.

In addition to these structural changes, functional connectivity changes have also been documented in chronic pain patients. However, to the best of the author's knowledge, there is currently no research or scientific evidence that relates the changes in functional connectivity to CRPS-I. In chronic back pain patients the insular cortex has increased co-activation with the Default Mode Network (Buckner, Andrews-Hanna, & Schacter, 2008) (comprised of: orbital middle frontal gyrus, left and right angular gyri and precuneus), indicating abnormal interactions between resting-state networks (Tagliazucchi, Balenzuela, Fraiman, & Chialvo, 2010). Furthermore, chronic patients show more frequent temporal fluctuations in the insular cortex and anterior cingulate cortex compared to controls (Malinen et al., 2010), demonstrating the abnormal changes in both the spatial and temporal characteristics of resting state networks in patients suffering from chronic pain. These cortical changes worsen as the syndrome persists, and often return to normal following successful treatment (Schweinhardt & Bushnell, 2010).

### **The 'Connectivity' Approach**

Previous research has already shown that neural activity can be indirectly captured using blood oxygen level dependent functional MRI (BOLD fMRI; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001), and in a resting-state study this signal is obtained in the absence of stimuli. Participants are often asked to relax with their eyes open or closed, or passively view a fixation cross. The aim of resting-state data analyses is to extract, from the raw data, functional networks of cortical regions that co-activate in the absence of stimulation (e.g. the Default Mode Network, see Buckner et al., 2008). Different approaches to this extraction have been developed, and these can be classified into two categories: confirmatory and exploratory. In confirmatory analyses, (such as seed-based techniques) a region-of-interest (ROI, or time-course-of-interest) is already known and is used to find regions that show spatiotemporal synchronicity. These regions are subsequently interpreted as forming a functional network with the region-of-interest (Huettel, Song, & McCarthy, 2004). In exploratory analyses (such as independent component

analysis, ICA), the raw data is decomposed into a set of maximally independent components – spatial or temporal – where knowledge of one component provides no information about another (Hyvärinen & Oja, 2000). These components are then interpreted as forming a functional network. A benefit of this approach is that no prior regions-of-interest need to be included in the analysis, hence its suitability for exploratory analyses (Calhoun, Adali, Hansen, Larsen, & Pekar, 2003). As previous research fails to provide enough robust evidence to justify a confirmatory analysis approach, ICA appears best-suited to the present study.

It should be noted that functional connectivity maps cannot be used to infer direct anatomical connectivity, as often there is no direct anatomical connectivity between regions with strong functional connectivity (Honey et al., 2009). In order to investigate anatomical connectivity *in vivo*, diffusion-weighted MRI (DW-MRI) techniques must be employed. DW-MRI captures the self-diffusion (Brownian motion) of water molecules, exploiting the phenomenon that in white matter fibre bundles, water diffusion is anisotropic [i.e. follows one direction more easily than others (Conturo et al., 1999)]. The DW-MRI data are used to describe the movement of water molecules using a diffusion tensor (symmetric 3x3 matrix). This diffusion tensor also allows analyses such as fibre-tracking (Basser, Pajevic, Pierpaoli, Duda, & Aldroubi, 2000), or fractional anisotropy (FA). FA is a voxel-wise index describing “...the fraction of the tensor that can be assigned to anisotropic diffusion” (Jones, 2008, p.940), that is, the directionality of intravoxel diffusion. Values range from 1 (anisotropic: diffusion along one direction) to 0 (isotropic: diffusion direction is equal in all directions) and this value is often interpreted as an amalgamation of the microstructural properties of white matter, and intravoxel boundary orientation (Beaulieu, 2002; Chen, Blankstein, Diamant, & Davis, 2011; Mukherjee et al., 2008).

For the purposes of studying a chronic pain population, resting-state fMRI and DW-MRI have a number of benefits over traditional fMRI methods. First, as patients often feel pain in the absence of stimulation, an investigation of neural activity at rest can provide insights into key differences between patients and controls. Second, the processing of pain is thought to be distributed across a number of cortical regions and not localised to a specific region, therefore a connectivity approach may be more appropriate. Third, the combination of structural and functional connectivity allows the exploration of the relationship between functional and structural connectivity.

## Research Questions and Hypotheses

The present study aims to investigate functional and anatomical connectivity in CRPS-I, using resting-state fMRI and DW-MRI, specifically focussing on changes in resting-state networks and fractional anisotropy. As of yet, this approach has not been used to study neuroplasticity in CRPS-I patients, and therefore we hope to elucidate: whether there are differences in connectivity between CRPS-I patients, and whether this novel methodology is an effective approach for studying chronic pain. This study will be divided into three research questions:

1. Are there differences in the spatio-temporal structure of *resting-state* networks between CRPS-I patients and controls?

2. Are there differences in fractional anisotropy between CRPS-I patients and controls?
3. Are fractional anisotropy and resting-state functional connectivity related?

Findings from previous research have led to the designation of ventromedial prefrontal cortex, insular cortex, anterior cingulate cortex, nucleus accumbens, primary and secondary somatosensory cortex as regions-of-interest. Tracts-of-interest include the cingulum bundle and inferior fronto-occipital fasciculus.

## METHOD

### Participants

Six participants (three controls, three patients) took part in this study. The average age was 45.2 years, with a range from 24 to 61 years. Participant characteristics are summarised below (table 1). The data from ‘study 1’ and ‘study 2’ were collected in the context of two larger studies, and therefore there are some variations in data acquisition parameters, and procedure (see ‘image acquisition’ and ‘procedure’ for more details). Only details relevant to the present study are described.

**Table 1.** Participant characteristics.

Participant	Gender	Age	Handedness	Affected hand	Current pain rating
Control 1 (study 1)	Female	55	Right	n/a	n/a
Control 2 (study 1)	Female	24	Left	n/a	n/a
Control 3 (study 2)	Male	26	Right	n/a	n/a
Patient 1 (study 1)	Female	61	Ambidextrous	Left	5
Patient 2 (study 1)	Female	55	Right	Right	2
Patient 3 (study 1)	Male	31	Right	Right	7.5

### Procedure

The present study required three MRI scans: resting-state fMRI, anatomical MRI, and diffusion-weighted MRI. In the scanner participants were in a supine position (head first), and instructed to remain as still as possible, there were no other visual or auditory cues. During the resting-state image acquisition participants were instructed to relax but remain awake with eyes closed (study 1), or keep their eyes open (study 2). Participants were provided with ear plugs, noise cancelling headphones with microphone and an emergency button. Before entering the scanner patients were asked to rate their current pain severity on a scale ranging

from 0 to 10 (0 = not at all, 10 = most pain ever felt).

### Image Acquisition

MRI data were acquired using a 3.0 Tesla Siemens Allegra scanner at Maastricht University, with an 8-channel head coil. BOLD (blood oxygen level dependent) signal was recorded using a T2\*-weighted gradient-echo echo planar imaging sequence. A total of 32 slices (voxel size of 3mm<sup>3</sup>, interleaved acquisition) were acquired [FOV = 216mm<sup>2</sup>, repetition time (TR), 2000ms; echo time (TE), 30ms; flip angle = 90°]. Diffusion-weighted images were obtained using a double re-focussed spin-echo EPI sequence (70 interleaved slices, voxel size of 2mm<sup>3</sup>, field-of-view 208mm<sup>2</sup>). Time-to-repeat (TR) was 8200ms, echo time (TE) 78ms and two b-values (0 s/mm<sup>2</sup> and 1000 s/mm<sup>2</sup>) were organised into 36 non-collinear directions (twelve b=1000 s/mm<sup>2</sup> to every one b=0 s/mm<sup>2</sup>). In study 1 a structural image was obtained using a T1-weighted sequence (ADNI-MPRAGE) with time-to-repeat (TR) 2250ms, time-to-echo (TE) 2.6ms and flip angle 9°. A total of 192 sagittal slices were obtained with a matrix size of 256mm<sup>2</sup> and an in-plane-resolution of 1mm<sup>3</sup>. In study 2 the structural image was obtained using a GRAPPA 2 sequence with voxel size of 1mm<sup>3</sup> (field-of-view 256mm<sup>2</sup>), time-to-repeat (TR) 2250ms, echo time (TE) 2.6ms and a flip angle of 9°.

### Data Analysis

#### *Resting-state fMRI*

Data were analysed using BrainVoyager QX (Brain Innovation, Maastricht, the Netherlands; Goebel, Esposito, & Formisano, 2006). The anatomical image was pre-processed to correct inhomogeneities in intensity and transformed into Talairach space using rotation into the ACPC plane and a piecewise scaling transformation. The automatic segmentation tool was used to segment the grey and white matter, disconnect the hemispheres and remove bridges in the reconstruction of the cortex.

A 3D reconstructed mesh was created from the grey/white matter segmentation—this was manually corrected for errors—then smoothed using 50 iterations and simplified to 80,000 vertices. This process was carried out per hemisphere and per participant. These meshes were used as the basis for cortex-based curvature alignment (Frost & Goebel, 2012). This involved creating a smoothed curvature map and morphing into a sphere. In order to facilitate multi-subject alignment these spheres were mapped to a standard sphere with a reduced number of vertices. The standard spheres were used for group alignment to a dynamic group average.

The resting-state data were then pre-processed to correct for the interleaved acquisition of slices (slice scan-time correction by sinc interpolation) and motion correction (trilinear/sinc interpolation). These pre-processed images were then co-registered to their respective anatomical images and used to create a volume time-course, which was transformed into Talairach space (trilinear interpolation with a target resolution of 3x3x3mm). Temporal filtering of the time-courses was applied using a high-pass filter (5 cycles) and a Gaussian kernel (FWHM = 4s). Spatial smoothing consisted of a gaussian kernel (FWHM = 6mm).

To remove physiological noise and residual motion artifacts ROIs were defined in the ventricles and white matter (intensity masking). The time courses from these ROIs were used as predictors in a single-study general linear model (z-transformed); the residuals from this analysis were saved as a new, corrected, volume time course.

This corrected time course was then projected onto the 3D mesh created for cortex-based curvature alignment, sampling -1mm and +3mm from each vertex. This was then used as the basis for cortex-based fast ICA, following dimensionality reduction to 30 components (1/6 of the number of time-points) using principal component analysis (Formisano, Esposito, Di Salle, & Goebel, 2004). The resulting independent components were then aligned onto the 'group average' so all participants' data were curvature aligned. The independent components were then clustered using cortex-based self-organising group ICA (Esposito et al., 2005).

Using the clusters which were created during the self-organising group ICA, a second level analysis (random effects analysis of variance) was used to test for the main effect of a subset of group clusters (each group cluster individually), and between-group differences within individual clusters ( $\alpha=0.05$ ). This subset of clusters was selected by first qualitatively assessing the validity of the clustering of the independent components, and secondly excluding group clusters that were characterised by spatial patterns localised to the temporal or occipital lobes. The regions identified by this test were then used as seed-ROIs for fibre-tracking (discussed below: 'tractography'). Regions that spread over multiple gyri and sulci were reduced in size by increasing the significance threshold. The ROIs used as seed-ROIs for fibre-tracking were also used for further analyses of functional connectivity. Each individual's time course, for each ROI, was extracted and analysed in MatLab. Correlation coefficients were calculated between each ROI (within a group-independent component) and transformed into z-scores using Fisher's transformation. A t-test was used to investigate between-group differences in functional connectivity for each tract. Following this, tracts which showed a significant difference in functional connectivity between patients and controls, and also a significant difference in FA (see below) were subject to a correlation analysis to test whether the changes in FA were related to changes in functional connectivity.

### *Tractography & Fractional Anisotropy*

ROIs were delineated on the cortex, using the results from the resting-state analysis. As the ROIs were identified from cortex-based aligned data, they were first transformed back into each individual participant's Talairach space, then to native space via ACPC. The ROIs were dilated (+1mm and -4mm from the selected vertex) so that they sufficiently overlapped with white matter.

The raw diffusion data was corrected for eddy-current distortion and motion in FSL (FMRIB, Oxford, United Kingdom; Smith et al., 2004; Woolrich et al., 2009) and then imported into BrainVoyager. Following this the diffusion image was co-registered to the anatomical image and voxels outside of the brain were masked. Sinc interpolation ( $R=3$ ) to a target resolution of 2x2x2mm was used to create a diffusion-weighted volume. From this diffusion-weighted volume the diffusion tensors were calculated, these tensors were used for fibre tracking and for the creation of FA maps.

Deterministic fibre tracking, step-size 0.5mm, (using diffusion tensor calculations from FA analysis) identified tracks connecting ROIs within a group cluster (FA threshold = 0.2; projection threshold = 0.25; angle threshold = 50). The FA values of the voxels within each tract were extracted and analysed using histogram tables (20 bins: FA values ranging from 0-0.05, 0.05-0.1, 0.1-0.15, etc.) which were calculated for the extracted FA values of each tract and each participant. The number of voxels per FA 'bin' were converted into percentages, and then an average percentage per FA 'bin' for patients and controls was calculated. An ordinal chi-squared test on each tract tested for a linear-by-linear association between group (patient versus control) and FA value (the 20 'bins'). Tracts which showed a significant linear-by-linear association, following Bonferroni's correction for multiple comparisons, were further analysed by comparing mean FA values, for patients and controls, using t-tests.

### *Fractional Anisotropy*

The raw diffusion-weighted data were corrected for eddy-current and motion distortions using affine registration to the first  $b=0$  s/mm<sup>2</sup> image [FDT toolbox (Behrens et al., 2003)]. Voxels outside of the brain were masked, and a binary brain was created [BET toolbox (Smith, 2002)]. The diffusion tensor model was fitted to the images and used to create FA maps [FDT toolbox (Behrens et al., 2003)]. The images were then fed into the tract-based spatial statistics pipeline [TBSS (Smith et al., 2006)] which began with non-linear registration [FNIRT (Andersson, Jenkinson, & Smith, 2007a, 2007b)] to align the images to a standard image (FMRIB FA image) (Rueckert et al., 1999). An average was made of all participants' FA images, this was then thinned by removing voxels below the threshold FA value (threshold FA = 0.2) to create a mean FA skeleton, upon which each participants' FA data were projected. This 4D image was used as the input for permutation based t-test analyses (patients versus controls), and the extraction of mean FA values (per participant) for ROIs (regions from the JHU white matter tractography atlas, Mori, Wakana, & van Zijl, 2005). These mean FA values were compared with t-tests (patients versus controls).

## RESULTS

### **Resting-state network analysis**

An ANOVA was used to test the main effect for each group cluster (i.e. identify the resting-state network), and a contrast to compare the between-group differences within each group cluster (i.e. differences in spatial coherence between patients and controls, for that resting-state network). The contrasts revealed significant differences in component spatial structure for patients versus controls, in all four group clusters ( $t_{(145)} > 3$ ,  $p = .0003$ , uncorrected for multiple comparisons, cluster-threshold = 50mm<sup>3</sup>).

Group cluster one consists of a RSN, from all participants, that demonstrates co-activation in bilateral inferior parietal cortex and right middle frontal gyrus. A small number of regions in right parietal and frontal cortex show differential coherence



within this network, for patients versus controls. Group cluster two consists of a RSN, from all participants, that demonstrates co-activation in bilateral posterior cingulate cortex as well as a diffuse region overlapping the precuneus and parieto-occipital sulcus. For this network, patients and controls showed differing network coherence in a small region in the precuneus and posterior cingulate cortex. The third group cluster was characterised by diffuse ventral frontal (including ventral medial) cortex activation, with group differences in network coherence in the left superior frontal sulcus. The fourth group cluster consists of a RSN, from all participants, that demonstrates co-activation in right superior frontal gyrus and a diffuse region spreading across bilateral inferior pre- and post-central gyri and the central gyrus. On the right hemisphere this network also extends into the operculum. Differences in network coherence, for patients versus controls, appeared in the bilateral inferior central sulcus.

The functional connectivity was calculated between the ROIs that were also connected by tracts identified using fibre-tracking (table 2). T-tests showed a significant difference in functional connectivity, between patients ( $M=0.66$ ,  $SD=0.07$ ) and controls ( $M=0.48$ ,  $SD=0.07$ ), for tract 9,  $t(4)=-2.893$ ,  $p=.04$  (uncorrected for multiple comparisons). This tract was defined as part of the right superior longitudinal fasciculus [JHU white matter atlas, (Mori et al., 2005)], connecting right supramarginal gyrus and post-central gyrus.

**Table 2.** Regions-of-interest and tracts-of-interest for the four group clusters. The regions identified by the resting-state ANOVA analysis, which are connected by white matter tracts, are labelled as regions-of-interest. Centre-of-Gravity, size in standard deviation (SD), and size in number of voxels is also included. Coordinates given are in Talairach space.

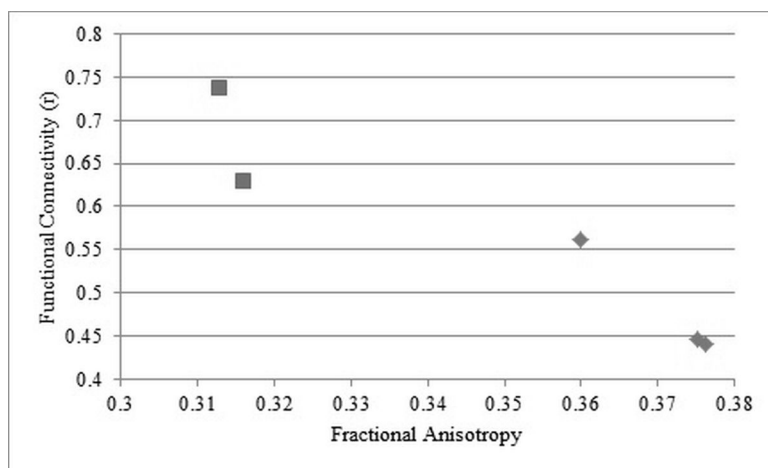
Region-of-Interest		Centre-of-Gravity			Size (Standard Deviation)			Size (Voxels)
		x	y	z	x	y	z	
Cluster 1: ROI 1	Right Inferior Parietal Lobule	35.01	-56.88	40.87	8.49	9.75	5.65	9432
Cluster 1: ROI 2	Right Inferior Frontal Gyrus	44.25	1.09	22.09	3.18	3.5	3.12	935
Tract 1: ROI 1 & ROI 2	Right Superior Longitudinal Fasciculus							
Cluster 2: ROI 1	Right Cingulate Gyrus	3.42	-29.13	27.69	2.18	5.79	4.11	1673
Cluster 2: ROI 2	Right Cuneus	10.12	-66.61	29.19	5.75	6.74	8.08	9546
Cluster 2: ROI 3	Left Precuneus	-13.47	-66.15	26.1	5.37	6.98	9.68	10764
Tract 2: ROI 1 & ROI 2	Forceps Major							
Tract 3: ROI 1 & ROI 3	Forceps Major							
Tract 4: ROI 2 & ROI 3	Forceps Major							
Cluster 3: ROI 1	Right Subcallosal Gyrus	12.32	24.62	-7.87	7.00	13.43	4.4	9906
Cluster 3: ROI 2	Left Subcallosal Gyrus	-11.19	16.24	-9.53	6.40	14.7	4.27	8282
Cluster 3: ROI 3	Right Inferior Frontal Gyrus	40.36	36.19	2.42	6.17	6.89	5.55	4197
Tract 5: ROI 1 & ROI 2	Forceps Minor							
Tract 6: ROI 1 & ROI 3	Forceps Minor							
Cluster 4: ROI 1	Right Precentral Gyrus	47.76	-13.12	34.21	5.18	5.04	6.95	5043
Cluster 4: ROI 2	Left Precentral Gyrus	51.22	-15.02	27.24	5.18	8.16	11.74	6529
Cluster 4: ROI 3	Left Precentral Gyrus	-47.16	-10.9	33.24	5.79	5.48	7.21	5635
Cluster 4: ROI 4	Right Postcentral Gyrus	-52.83	-11.49	15.36	2.03	3.07	2.53	576
Cluster 4: ROI 5	Inferior Parietal Lobule	51.37	-37.36	31.58	3.16	3.12	4.71	1430
Tract 7: ROI 1 & ROI 2	Right Superior Longitudinal Fasciculus							
Tract 8: ROI 1 & ROI 5	Right Superior Longitudinal Fasciculus							
Tract 9: ROI 2 & ROI 5	Right Superior Longitudinal Fasciculus							
Tract 10: ROI 4 & ROI 3	Left Superior Longitudinal Fasciculus							



## Tractography

Deterministic fibre-tracking identified a total of 10 tracts which connected ROIs within the components; these tracts were used as the basis for the comparison of FA values. The ordinal chi-square test revealed a significant linear-by-linear association between group and FA value for tract 1 ( $M^2=8.18$ ,  $df=1$ ,  $p=.04$ ), tract 8 ( $M^2=41.55$ ,  $df=1$ ,  $p=.02$ ), tract 9 ( $M^2=14.81$ ,  $df=1$ ,  $p<.001$ ), and tract 10 ( $M^2=9.15$ ,  $df=1$ ,  $p<.001$ ). Tracts 1, 8 and 9 are part of the right superior longitudinal fasciculus, whilst tract 10 is part of the left superior longitudinal fasciculus [JHU white matter atlas, (Mori et al., 2005)].

Tract 9 showed both a significant difference in functional connectivity, and FA values, for patients versus controls. A correlation analysis revealed a strong negative correlation between functional connectivity and FA in this tract,  $r=-0.90$ ,  $p=.02$  (figure 1).



**Figure 1:** Graph showing functional connectivity as a function of fractional anisotropy, in tract 9, for each participant (Square = Patients; Diamond = Controls). There is a significant negative correlation between these measures ( $r=-0.90$ ,  $p=.02$ ). This tract is part of the right superior longitudinal fasciculus, and connects the right supramarginal gyrus and post-central gyrus.

## Fractional Anisotropy

A voxel-wise t-test of the FA skeleton revealed no significant differences in FA values between patients and controls ( $\alpha=.05$ ). The mean FA values of the FA skeleton, left cingulum, right cingulum, and left inferior fronto-occipital fasciculus did not differ significantly different for patients versus controls. The difference between patients and controls for the right inferior fronto-occipital fasciculus was significant at  $p<.05$  (see table 3).

**Table 3.** Comparison of fractional anisotropy values for CRPS-I patients compared to controls. \*Significant difference between patients and controls ( $\alpha=.05$ ).

Tract-of-interest	Patient	Controls	T-test
Mean FA skeleton	$M=0.41, SD=0.01$	$M=0.43, SD=0.01$	$t(4)=2.66, p=.057$
Left Cingulum	$M=0.45, SD=0.02$	$M=0.47, SD=0.02$	$t(4)=2.02, p=.11$
Right Cingulum	$M=0.44, SD=0.03$	$M=0.46, SD=0.02$	$t(4)=1.28, p=.27$
Left Inferior Fronto-occipital Fasciculus	$M=0.42, SD=0.02$	$M=0.45, SD=0.01$	$t(4)=1.77, p=.15$
Right Inferior Fronto-occipital Fasciculus	$M=0.40, SD=0.01$	$M=0.43, SD=0.01$	$t(4)=3.68, p=.02^*$

## DISCUSSION

The three research questions of this study focussed on: differences in spatiotemporal structure of resting-state networks; differences in white matter integrity; and whether these differences are related, for CRPS-I patients and controls. The results demonstrate that there are differences in functional and anatomical connectivity between CRPS-I patients and healthy controls. Specifically, resting-state networks show differential spatial structure, as well as increased functional connectivity between the right supramarginal gyrus and post-central gyrus, for CRPS-I patients compared to controls. Furthermore, portions of the right superior longitudinal fasciculus, which connected these regions, were found to have lower fractional anisotropy values for CRPS-I patients. Interestingly, the lower a participant's fractional anisotropy value, the higher their functional connectivity between these regions.

With regards to the differences in resting-state network spatiotemporal structure between controls and patients, the results support the hypotheses that resting-state networks show differential spatial structure and functional connectivity. Of particular functional relevance is the network referred to as 'group cluster 4', which appears to be consistent with a sensorimotor resting-state network (Biswal, Yetkin, Haughton, & Hyde, 1995; De Luca, Beckmann, De Stefano, Matthews, & Smith, 2006). The sensorimotor resting-state network is particularly relevant for CRPS-I, as the condition is characterised by a range of sensory and motor symptoms (e.g. hyperalgesia, bradykinesia). Within this network, the functional connectivity between the right supramarginal gyrus and post-central

gyrus was stronger for CRPS-I patients compared to controls. The post-central gyrus corresponds with the functional region of primary somatosensory cortex, whilst the region in the supramarginal gyrus appears to be consistent with the functional region of secondary somatosensory cortex (Eickhoff, Grefkes, Fink, & Zilles, 2008; Eickhoff, Amunts, Mohlberg, & Zilles, 2006). Both of these regions are also part of the sensory-discriminative network of pain processing (Peyron, Laurent, & García-Larrea, 2000).

This increased functional connectivity between right primary and secondary somatosensory cortex was ipsilateral to the affected limb for two patients and contralateral to the affected limb for one patient. As the data were not aligned by affected limb, nor were there enough patients to test the effect of the side of the affected limb, it is not possible to draw a robust conclusion from this result. Possible interpretations of this result are: (i) the altered functional connectivity is related to the side of the affected limb (perhaps the result of compensatory behaviours), (ii) the altered functional connectivity is lateralised to the right hemisphere, irrespective of the side of the affected limb. This is similar to the result observed by Geha et al. (2009), that in CRPS-I patients there was a lateralised decrease in grey matter density of the right anterior insular cortex.

Concerning the anatomical connectivity, the white matter integrity of the superior longitudinal fasciculus (subdivision III), which connects the supramarginal gyrus with the ventral pre-motor and pre-frontal regions, differed significantly for CRPS-I patients compared to controls. Specifically, fractional anisotropy was lower for patients than controls in the left portion of the superior longitudinal fasciculus, and the distribution of values for both the right and left portions was different. These results strongly suggest that CRPS-I patients have altered white matter integrity in this—perhaps even a global decrease in white matter integrity. Consistent with this conclusion is the result that there is a trend for lower fractional anisotropy values in both bilateral cingulum and bilateral inferior fronto-occipital fasciculus, as well as a decrease in overall mean fractional anisotropy ( $p=.058$ ). The pattern of lowered fractional anisotropy values is also consistent with the results from the study by Geha et al. (2009), where they found decreased fractional anisotropy in the left cingulum-callosal bundle.

Finally, the third research question was directed at the relationship between fractional anisotropy and functional connectivity. The functional connectivity of the third subdivision of the right superior longitudinal fasciculus correlated negatively with the FA values. If lower fractional anisotropy values indicate compromised white matter integrity, then an indirect white matter pathway may modulate the functional connectivity. However, if the decrease in fractional anisotropy is due to a factor which improves the propagation of action potentials, then this white matter tract may facilitate the functional connectivity in CRPS-I patients. Since it is not possible to conclude that lower fractional anisotropy values indicate damaged white matter, it is not possible to draw a concrete conclusion from this result (Jones, Knösche, & Turner, 2012).

The total number of participants limits the present study: with a larger sample the statistical power would increase and the analysis could take into account the large degree of intra-group variability. Patients with CRPS-I often suffer from

a wide range of symptoms, and as a result a dichotomous patient and control comparison may not be accurate. Pain duration and severity, which have previously been shown to exacerbate cortical changes, further confound the assumption that CRPS-I patients can be grouped as a homogenous population (Baliki, Schnitzer, Bauer, & Apkarian, 2011b; Chen et al., 2011; Geha et al., 2009). Therefore, a larger sample of participants would allow the introduction of a within-group variable of symptom, and also pain duration and/or severity. These changes would hopefully increase the sensitivity of the analyses to cortical changes associated with CRPS-I, specific symptoms, and the effect of pain duration and severity. Additionally, the matching of controls and patients could be improved in order to reduce the effect of demographic variability, and patients' data could be aligned by affected limb to avoid variability due to lateralisation of cortical changes.

With regards to image acquisition, participants in study one were instructed to keep their eyes closed during the resting-state scan. However, research has demonstrated that resting-state networks show significantly stronger correlations when participants are instructed to keep their eyes open and focussed on a fixation cross (Van Dijk et al., 2010). The (control) participant in study two was instructed to keep his eyes open during the resting-state scan, although no fixation cross was presented due to practical limitations. It is unlikely that the different instruction given to study one and study two participants would have a significant effect on the resting-state data, as the study by van Dijk et al. (2010) found no significant difference between eyes open and eyes closed, but if the present study was replicated then it would be advisable to change the instruction to passive fixation.

In the resting-state data analysis there was a conversion from a whole brain to cortex-based data set (Formisano et al., 2004). This step has two main benefits, which are: a) the subsequent cortex-based alignment for group analyses, and b) the limiting of ICA to grey matter voxels. This is especially useful for ICA as one of the key assumptions underlying most ICA approaches is that the components are mixed in a linear fashion. McKeown and Sejnowski (1998) assessed the validity of these assumptions and found that non-linear mixing models may be more appropriate than linear mixing models. On the other hand, if the data is restricted to grey matter voxels, then the assumption of linear mixing is still valid (McKeown & Sejnowski, 1998). Furthermore, cortex-based alignment has been shown to improve inter-subject alignment, especially in frontal regions and regions around the central sulcus—both of which were regions-of-interest in the present study (Frost & Goebel, 2012). However, there is a risk involved with summarising grey matter data onto the vertices of the cortex mesh: depending on the grey and white matter segmentation, the data represented on the vertex may be sampled too deep (into white matter) and therefore not only summarising grey matter voxels.

The group level resting-state data analysis used the self-organising ICA approach. One limitation of this approach is that when the individual independent components are clustered based on spatial similarity, then there are not (by definition) *large* between-group differences in resting-state network spatial structure. Whilst this does limit comparisons of the spatial structure of a network, this aspect of self-organising ICA does not affect the functional connectivity calculations. For the purposes of this study this benefit outweighed the limitation,

as the functional connectivity correlations are the quantification of the strength and directionality (i.e. negative or positive) of the connectivity between regions.

To investigate differences in anatomical connectivity, the present study used a deterministic fibre-tracking method, which does not give an indication of how reliable the tracking result is. Within each voxel there is a degree of error associated with the determined fibre orientation, a fact exploited by probabilistic fibre-tracking methods. These approaches incorporate the probability density function of fibre orientations within a voxel, resulting in a map which quantifies the degree of certainty associated with the tract (Jones, 2008). Whilst this is a clear advantage over the deterministic approach, probabilistic tracking can (usually) only be carried out from one seed-point and not a multi-voxel region-of-interest and therefore was not a feasible approach for this study.

Two approaches to comparing fractional anisotropy values were used in this study: the first used tracts-of-interest defined using a white matter atlas, the second using fibre-tracking from functionally defined regions-of-interest. Whilst the atlas approach has a number of benefits in terms of anatomically focussed research questions, the functional approach integrates both functional and anatomical data (Kanaan et al., 2006). Essentially, both approaches are attempting to ensure that when comparing two or more participants, the data is well aligned. In the skeleton and mask approach the data is structurally aligned, so in principle location  $x$  corresponds to the same structure in all participants. In the fibre-tracking approach the data is functionally aligned: that is, the white matter being compared connects the same functionally-relevant grey matter regions. The advantage of one approach over another is largely dependent on the research question being asked. For example, is there a specific location or structure of interest? If so, then the skeleton and mask approach will allow a structural alignment of data and a 'standardisation', so that there are an equal number of voxels per person, per tract (a benefit for a number of statistical analyses). If the researcher is interested in the white matter that connects to (e.g.) insular cortex, then the fibre-tracking approach may be more relevant. This captures individual differences in the white matter connected to a specific region, and therefore, may be more useful for gauging the interaction between white matter, and its connected grey matter. However, there are some inherent limitations with both approaches: in the skeleton and mask approach there is an assumption that the white matter of the participants is well aligned, whilst no alignment is needed in the fibre-tracking approach, there is an assumption that the regions-of-interest are similarly placed for each participant. In the present study the regions-of-interest were demarcated on a group-average curvature-aligned cortex mesh, and then aligned to the participant's native space, via a number of transformations. However, one of these transformations could have introduced a large degree of variability: the projection of the region-of-interest from the cortex mesh onto a whole-brain volume (i.e. from a vertex to grey matter, which can vary in thickness). A consistent approach was adopted for this (+1mm and -4mm); however, the accuracy of this projection may have varied from participant to participant. Differences in grey matter thickness, or slight variations in the transformation to native space, may have resulted in varying degrees of overlap with white matter and thus affected the fibre-tracking results.

The present study questions whether or not functional and anatomical connectivity differs between CRPS-I patients and controls, and whether these differences relate to each other. The results suggest that functional connectivity increases between primary and secondary somatosensory cortex (both of which are part of the sensorimotor resting-state network) for CRPS-I patients. Interestingly, the white matter connecting these regions showed significantly decreased fractional anisotropy, and this correlates negatively with the functional connectivity observed. Whilst this study had a number of methodological limitations which prevent the drawing of robust conclusions, it demonstrates the feasibility of applying this analysis pipeline to patients with chronic pain. A number of amendments have been suggested to improve this analytical approach, which could allow a more comprehensive and sensitive investigation of neuronal connectivity in chronic pain.

## REFERENCES

- Alexander, G. M., Peterlin, B. L., Perreault, M. J., Grothusen, J. R., & Schwartzman, R. J. (2011). Changes in Plasma Cytokines and Their Soluble Receptors in Complex Regional Pain Syndrome. *The journal of pain : official journal of the American Pain Society*, 13(1), 10–20. doi:10.1016/j.jpain.2011.10.003
- Andersson, J., Jenkinson, M., & Smith, S. (2007a). *Non-linear optimisation*. Retrieved from [www.fmrib.ox.ac.uk/analysis/techrep](http://www.fmrib.ox.ac.uk/analysis/techrep)
- Andersson, J., Jenkinson, M., & Smith, S. (2007b). *Non-linear registration, aka spatial normalisation*. Retrieved from [www.fmrib.ox.ac.uk/analysis/techrep](http://www.fmrib.ox.ac.uk/analysis/techrep)
- Baliki, M. N., Schnitzer, T. J., Bauer, W. R., & Apkarian, a. V. (2011a). Brain Morphological Signatures for Chronic Pain. (R. M. Luque, Ed.) *PLoS ONE*, 6(10), e26010. doi:10.1371/journal.pone.0026010
- Baliki, M. N., Schnitzer, T. J., Bauer, W. R., & Apkarian, a. V. (2011b). Brain Morphological Signatures for Chronic Pain. (R. M. Luque, Ed.) *PLoS ONE*, 6(10), e26010. doi:10.1371/journal.pone.0026010
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, a. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 44(4), 625–32. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11025519>
- Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in biomedicine*, 15(7-8), 435–55. doi:10.1002/nbm.782
- Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic resonance in medicine*, 50(5), 1077–1088.
- Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. *Magnetic resonance in medicine*, 34, 537–541.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38. doi:10.1196/annals.1440.011
- Calhoun, V. D., Adali, T., Hansen, L. K., Larsen, J., & Pekar, J. J. (2003). ICA of functional MRI data: An overview. *Framework*, 281–288.
- Chen, J. Y., Blankstein, U., Diamant, N. E., & Davis, K. D. (2011). White matter abnormalities in irritable bowel syndrome and relation to individual factors. *Brain research*, 1392, 121–131.



- doi:10.1016/j.brainres.2011.03.069
- Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., et al. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 96(August), 10422–10427.
- De Boer, R. D. H., Marinus, J., Van Hilten, J. J., Huygen, F. J., Van Eijs, F., Van Kleef, M., Bauer, M. C. R., et al. (2011). Distribution of signs and symptoms of Complex Regional Pain Syndrome type I in patients meeting the diagnostic criteria of the International Association for the Study of Pain. *European journal of pain*, 15(8), 830.e1–830.e8. doi:10.1016/j.ejpain.2011.01.012
- De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., & Smith, S. M. (2006). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *NeuroImage*, 29(4), 1359–67. doi:10.1016/j.neuroimage.2005.08.035
- De Mos, M., De Bruijn, G. J., Huygen, F. J. P. M., Dieleman, J. P., Stricker, B. H. C., & Sturkenboom, M. C. J. M. (2007). The incidence of complex regional pain syndrome: a population-based study. *Pain*, 129(1-2), 12–20. doi:10.1016/j.pain.2006.09.008
- Eickhoff, S. B., Grefkes, C., Fink, G. R., & Zilles, K. (2008). Functional lateralization of face, hand, and trunk representation in anatomically defined human somatosensory areas. *Cerebral cortex (New York, N.Y. : 1991)*, 18(12), 2820–30. doi:10.1093/cercor/bhn039
- Eickhoff, Simon B., Amunts, K., Mohlberg, H., & Zilles, K. (2006). The human parietal operculum. II. Stereotaxic maps and correlation with functional imaging results. *Cerebral cortex (New York, N.Y. : 1991)*, 16(2), 268–79. doi:10.1093/cercor/bhi106
- Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., Tedeschi, G., et al. (2005). Independent component analysis of fMRI group studies by self-organizing clustering. *NeuroImage*, 25(1), 193–205. doi:10.1016/j.neuroimage.2004.10.042
- Formisano, E., Esposito, F., Di Salle, F., & Goebel, R. (2004). Cortex-based independent component analysis of fMRI time series. *Magnetic resonance imaging*, 22(10), 1493–504. doi:10.1016/j.mri.2004.10.020
- Frost, M. a., & Goebel, R. (2012). Measuring structural-functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, 59(2), 1369–81. doi:10.1016/j.neuroimage.2011.08.035
- Geha, P. Y., Baliki, M. N., Harden, R. N., Bauer, W. R., Todd, B., & Apkarian, A. V. (2009). The brain in chronic CRPS pain: Abnormal gray-white matter interactions in emotional and autonomic regions. *Neuron*, 60(4), 570–581. doi:10.1016/j.neuron.2008.08.022.
- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, 27, 392–401.
- Harke, H., Grentenkort, P., Ladleif, H. U., Rahman, S., & Harke, O. (2001). The response of neuropathic pain and pain in complex regional pain syndrome I to carbamazepine and sustained-release morphine in patients pretreated with spinal cord stimulation: A double-blinded randomised study. *Pain*, (9), 488–495.
- Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), 2035–40. doi:10.1073/pnas.0811168106
- Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional Magnetic Resonance Imaging. Magnetic Resonance Imaging* (1st Editio.). Sunderland, MA: Sinauer Associates.
- Hyvärinen, a., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks : the official journal of the International Neural Network Society*, 13(4-5), 411–30. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10946390>
- Iannetti, G. D., & Mouraux, A. (2010). From the neuromatrix to the pain matrix (and back).



- Experimental brain research*, 205(1), 1–12. doi:10.1007/s00221-010-2340-1
- Jones, D. K. (2008). Studying connections in the living human brain with diffusion MRI. *Cortex; a journal devoted to the study of the nervous system and behavior*, 44(8), 936–52. doi:10.1016/j.cortex.2008.05.002
- Jones, D. K., Knösche, T. R., & Turner, R. (2012). White Matter Integrity, Fiber Count, and Other Fallacies: The Do's and Don'ts of Diffusion MRI. *NeuroImage*, 1–16. doi:10.1016/j.neuroimage.2012.06.081
- Kanaan, R. a, Shergill, S. S., Barker, G. J., Catani, M., Ng, V. W., Howard, R., McGuire, P. K., et al. (2006). Tract-specific anisotropy measurements in diffusion tensor imaging. *Psychiatry research*, 146(1), 73–82. doi:10.1016/j.psychres.2005.11.002
- Kasper, D. L., Braunwald, E., Fauci, A. S., Hauser, S. L., Longo, D. L., & Jameson, J. L. (2005). *Harrison's Principles of Internal Medicine* (16th ed., p. 1907). McGraw-Hill.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, a. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–7. doi:10.1038/35084005
- Maihöfner, C., Handwerker, H. O., & Birklein, F. (2006). Functional imaging of allodynia in complex regional pain syndrome. *Neurology*, 66(5), 711–7. doi:10.1212/01.wnl.0000200961.49114.39
- Maihöfner, C., Neundörfer, B., Birklein, F., & Handwerker, H. O. (2006). Mislocalization of tactile stimulation in patients with complex regional pain syndrome. *Journal of neurology*, 253(6), 772–9. doi:10.1007/s00415-006-0117-z
- Malinen, S., Vartiainen, N., Hlushchuk, Y., Koskinen, M., Ramkumar, P., Forss, N., Kalso, E., et al. (2010). Aberrant temporal and spatial brain activity during rest in patients with chronic pain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14), 6493–7. doi:10.1073/pnas.1001504107
- Marinus, J., Moseley, G. L., Birklein, F., Baron, R., Maihofner, C., Kingery, W. S., & Van Hilten, J. J. (2011). Clinical features and pathophysiology of complex regional pain syndrome. *Lancet neurology*, 10(7), 637–648. doi:10.1016/S1474-4422(11)70106-5
- McKeown, M J, Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, a J., & Sejnowski, T. J. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human brain mapping*, 6(3), 160–88. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9673671>
- McKeown, Martin J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, 6(5-6), 368–372. doi:10.1002/(SICI)1097-0193(1998)6:5/6<368::AID-HBM7>3.0.CO;2-E
- Mori, S., Wakana, S., & Van Zijl, P. (2005). *Atlas of Human White Matter*. Amsterdam, the Netherlands: Elsevier.
- Mukherjee, P., Berman, J. I., Chung, S. W., Hess, C. P., & Henry, R. G. (2008). Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *American journal of neuroradiology*, 29(4), 632–41. doi:10.3174/ajnr.A1051
- Patterson, R. W., Li, Z., Smith, B. P., Smith, T. L., & Koman, L. A. (2011). Complex regional pain syndrome of the upper extremity. *The Journal of hand surgery*, 36(9), 1553–62. doi:10.1016/j.jhsa.2011.06.027
- Peltz, E., Seifert, F., DeCol, R., Dörfler, A., Schwab, S., & Maihöfner, C. (2011). Functional connectivity of the human insular cortex during noxious and innocuous thermal stimulation. *NeuroImage*, 54(2), 1324–35. doi:10.1016/j.neuroimage.2010.09.012
- Peyron, R., Laurent, B., & García-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Clinical neurophysiology*, 30(5), 263–88. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11126640>
- Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., & Hawkes, D. (1999). Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE transactions on medical imaging*, 18(8), 712–721.

- Schweinhardt, P., & Bushnell, M. C. (2010). Pain imaging in health and disease — how far have we come ? *The Journal of clinical investigation*, 120(11), 3788–3797. doi:10.1172/JCI43498.3788
- Smith, S. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3), 143–155.
- Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T., Mackay, C., Watkins, K., et al. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31, 1487–1505.
- Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., et al. (2004). Advances in functional and structural MRI image analysis and implementation as FSL. *NeuroImage*, 23(1), 208–219.
- Stejskal, E. O., & Tanner, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42(1), 288–292.
- Swart, C. M. A., Stins, J. F., & Beek, P. J. (2009). Cortical changes in complex regional pain syndrome (CRPS). *European journal of pain*, 13(9), 902–907. doi:10.1016/j.ejpain.2008.11.010
- Tagliazucchi, E., Balenzuela, P., Fraiman, D., & Chialvo, D. R. (2010). *Brain resting state is disrupted in chronic back pain patients. Neuroscience letters* (Vol. 485, pp. 26–31). doi:10.1016/j.neulet.2010.08.053
- Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, 103(1), 297–321. doi:10.1152/jn.00783.2009
- Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45, S173–186.

# Cabbage, Cars and Beer? An examination of Dual Attitudes towards the Concept “German”

## ORIGINAL PAPER

A controversial and still ongoing debate concerns the question whether observed weak correlations between implicit and explicit attitudes point to two different underlying constructs of attitudes or emerge due to assessing the same construct in different ways. Next to addressing the problem of the low explanatory power of correlational conclusions, the present study aimed to contribute behavioral evidence to this debate with an additional focus to the underrepresented single target categories in implicit social cognition, such as nationality concepts. The results of the assessment of attitudes towards the concept “German” point to the conclusion that in general, attitudes towards the concept are mostly negative. Implications from these results are used to develop strategies regarding the facilitation of academic student mobility.

**Keywords:** attitudinal dissociation, nationality concepts, academic student mobility

Sabine Lenz; Master student  
Maastricht University, Maastricht, the Netherlands

---

s.lenz@student.maastrichtuniversity.nl

“Attitude is a little thing that makes a big difference.” —Winston Churchill

## INTRODUCTION

Although Churchill was most likely referring to a rather colloquial use of the word attitude he holds some point. Our attitudes, defined as “a summary evaluation of

an object of thought” (Bohner & Wänke, 2002, p. 5) have been a controversial topic since the early years of social psychology. A very wide range of particular behaviors and beliefs can be compiled under the label of one specific attitude (Oskamp & Schultz, 2005) causing persisting confusion what to put under the label of “attitude”.

Nevertheless, attitudes are clearly influential. We socialize easier with people that hold comparable attitudes towards topics we are concerned with, and withdraw from people and/or situations that are not matching our own beliefs (Bohner & Wänke, 2002). The main reason for studying attitudes is therefore certainly the assumption that attitudes influence a person’s behavior and can be used to explain and predict behavior. However, this might only be correct at first glance. Real life shows us that we often act contrary to our attitudinal beliefs (Bohner & Wänke, 2002). For instance, there is only surprisingly little evidence for a strong relation between students’ attitude towards cheating on exams and their actual behavior of cheating (Wicker, 1969). Thus, it is rather the question in *how far* our behavior is influenced by attitudes and whether the attitude shapes behavior or vice versa. Probably, behavior and attitudes should be even considered as being reciprocally active (Oskamp & Schultz, 2005).

### **A brief history of attitude research and its problems**

In the past, social psychological research was primarily focused on the nature of explicit attitudes and their predictive power for behavior (e.g., Fishbein & Ajzen, 1977, 2005). These studies explored when and under which circumstances explicit attitudes could be used for anticipating behavior. Much to the dismay of the researchers it was found that a single, isolated attitude will most likely not predict distinct behaviors. Instead, attitudes give us some indication of an average behavioral tendency. Hence, attitudes *can* contribute to explanatory and predictive features of behavior, but certain frame conditions (e.g., minimization of other influences to the behavior, potency of attitude et cetera) have to be met. In addition, another disadvantage of explicit attitude measurement became prevalent: the vulnerability of explicit attitude measurement to socially desirable answering tendencies (e.g., Hofmann, Gschwendner, Castelli, & Schmitt, 2008). Due to the fact that explicit attitudes are assessed with self-reports in most cases, people are prone to adjust their reports towards socially more accepted positions (Greenwald, McGhee, & Schwartz, 1998).

Consequently it is the concept of implicit attitudes and its measurement, as well as its relation to explicit attitudes that engages social psychological research over the past 15 years (e.g., Greenwald & Banaji, 1995; Wilson, Schooler & Lindsey, 2000; Hofmann et al., 2008).

### **Implicit attitudes – a solution causes problems**

Implicit attitudes are by definition positive or negative emotions, thoughts, or behaviors towards social objects which are affected by past experience. This past experience, however, remains either completely inaccessible to the person or is only imprecisely recalled (Greenwald & Banaji, 1995). Therefore, implicit attitudes are much less susceptible in their measurement to trigger biases such as social desirable

answering tendencies (Greenwald et al., 1998). Accordingly, new measurement techniques of implicit attitudes have been developed (for a summary see e.g., Fazio & Olson, 2003) to circumvent the problem of measuring attitudes explicitly, and an enormous amount of research employing implicit measurement techniques has been conducted (Hofmann et al., 2008).

However, the concept of implicit attitude is still called into question with emphasis on its general existence, nature and measurement. This is due to the fact that behavioral research sometimes uncovers astonishing results with regard to people's implicit and explicit attitudes. It seems to be the case that especially the evaluation of social controversial topics, such as racial prejudice or sexual orientation, elicits a phenomenon that Greenwald and Banaji (1995) referred to as a "dissociation of attitudes". This dissociation is reflected in the weak correlations between implicit and explicit measures that are observed. People seem to have two different attitudes towards the same topic (e.g., Monteith, Voils, & Ashburn-Nardo, 2001; Dovidio, Kawakami, & Gaertner, 2002). For instance, some people who claim to have an open and agreeable attitude towards homosexuality show response patterns on implicit attitude testing that are strongly indicative of a negative attitude towards same-sex orientation (e.g., Jellison, McConnel, & Gabriel, 2004; Steffens, 2005). The precise reasons for this surprising phenomenon are still unclear.

### **Theories on attitudinal dissociation**

Since the initial proposition by Greenwald and Banaji (1995) that there is a whole category of attitudes which is inaccessible by self-reports and the like, because of its unconscious mode of processing, numerous theories about the conceptual nature of the implicit/explicit attitude relationship were constructed (Nosek & Smyth, 2007). Most of the theories about the nature of implicit attitudes thereby focus on approaches to *explain* the differences between implicit and explicit attitudes in behavioral data. The major interpretations that contributed to the debate of the suspected two types of attitudes can be summarized into three broad categories (Greenwald & Nosek, 2009): (a) *single-representation interpretations* - This type of interpretation addresses the phenomenon of attitudinal dissociation as an illusion. The difference between the outcomes of explicit and implicit measures of the same attitude object is caused by assessing this attitude representation with different measurement techniques (e.g., Fazio & Olson, 2003); (b) *dual-representation interpretations* - This viewpoint assumes that attitudinal dissociation arises from the separated, conceptual structure of the underlying attitudes tapped. More specifically, explicit attitudes result from a conscious and deliberate process, whereas implicit attitudes arise from an automatic and probably unconscious mode (Wilson et al., 2000; Strack & Deutsch, 2004); (c) *Person vs. culture interpretations* - The last major type of interpretation assumes that the attitudinal dissociation originates in two different kinds of influences a person is exposed to throughout life: The culture in which one is living, and the personal development. Thereby, cultural frameworks of semantic knowledge affect implicit measures. In turn, explicit measures delineate personal influences (e.g., Karpinski & Hilton, 2001).

Whereas the third approach received relatively little attention in psychological

research, the first two approaches caused a persistent and still ongoing debate. Fazio and Olson (2003) who support the first approach, suggest the use of the so called “motivation and opportunity as determinants”- model (MODE Model) for the attitudinal dissociation phenomenon. In this model, the attitude-to-behavior process is never completely unconscious, nor entirely intentional. Instead, it is a mixed process featuring both components (Fazio & Ohlson, 2003). That is to say, different measurement techniques of implicit and explicit attitude explain the observed distinction between the two types of attitude, because the techniques measure different stages of processing of a single attitude object rather than tapping two distinct representations. In contrast, Wilson et al. (2000) developed a dual model of attitudes theorizing the finding that implicit attitudes seemed to be conceptually different from explicit attitudes in terms of attitudinal dissociation (Greenwald & Banaji, 1995). In this model, implicit and explicit attitudes are understood as independent, coexisting evaluations of an attitude object. People can have two attitudes towards the same object at the same time: implicit attitudes (usually processed unconsciously) and explicit attitudes (which are part of the consciously aware mindset). This viewpoint is supported by evidence from several studies as well as by other theoretical models. For instance, Strack and Deutsch (2004) proposed a two-system model that expands the dual attitude model of Wilson et al. (2000). They propose that the phenomenon of attitudinal dissociation can be explained in terms of tapping distinct evaluative sources. Thereby, explicit attitudes are attributed to a reflection of people’s opinions and are therefore the consequence of a relatively deliberate process of thought. Implicit attitudes, in turn, are thought to be activations of people’s associative structures, a rather unconscious mode of thought that is characterized by fast and non-normative processing. Additionally, evidence for the assumption that implicit and explicit attitudes are indeed two independent constructs can be obtained from a wide variety of research. Gawronski and Strack (2004), for example, showed that an induced cognitive dissonance affects explicit racial prejudice, but implicit prejudice remains unaffected. Similar patterns are found in other studies concerning social controversial topics, stating that implicit and explicit measures of racial attitudes as well as ethnocentrism are related but distinct factors (Cunningham, Preacher, & Banaji, 2001; Cunningham, Nezlek, & Banaji, 2004 as cited in Nosek & Smyth, 2007). Finally, probably the most important evidence with regard to the dual representation approach stems from a large-scale re-evaluation of implicit construct validity by Nosek and Smyth (2007). Following this re-evaluation, the constructs of implicit and explicit attitudes are in fact related, but the deeper analysis of data suggests two distinct underlying constructs of attitude. Thus all evidence stated above might be interpreted as indicative for a dual representational viewpoint of attitudes. However, the variability if and in how far implicit and explicit attitudes are connected cannot be fully explained by now. It is obvious that there is still a need for additional research to understand the nature of the relationship between implicit and explicit attitudes.

It becomes also more and more evident that the phenomenon of attitudinal dissociation could be at least partially attributed to certain methodological issues in the assessment of implicit attitudes. Most of the studies mentioned above made use of the Implicit Association Test (IAT) as a measurement of implicit attitudes

(e.g., Greenwald et al., 1998; Monteith et al., 2001; Cunningham et al., 2001). The IAT however has one major disadvantage which could attenuate the evidence used to examine possible conclusions. Due to its need of a contrasting category, it is only the relative strength of associations between target category and contrasting category that can be assessed (Greenwald et al., 1998). Next to the rather disconcerting idea that a methodological issue could be held responsible for a phenomenon which causes controversy in an entire field of psychology, some concepts in social psychology simply remain unexamined due to this fact. Considering that not all concepts of social implicit cognition have a contrasting category (like in a racial IAT with black and white skin color as target categories or with Pepsi and Coke), it is particularly problematic to examine such target concepts in general (Nosek & Banaji, 2001).

### **Single target evaluation – Just black and white are not enough**

An example for such a concept without a contrasting counter category would be the concept of “nationality”. This concept is usually defined as the belonging to a nation, most of the time influenced by the individual’s citizenship, but in some cases also by ethnicity or place of residence or the individual’s sense of national identity (“nationality,” Oxford Dictionary, 2000). Given its nature of being a rather fuzzy concept instead of a clear cut category, it is obvious that it cannot be evaluated with regard to counter categories. Consider for instance, the concept “German”. Some more or less loose associations like “Oktoberfest”, German cars, beer or the unavoidable Nazi Germany might come up (Bolten, 2006). Moreover, most people will have some kind of attitude towards the concept “German” as well. However, only little attention has been paid on attitudes towards nationality concepts due to the afore-mentioned problematic evaluation. Any the less, nationality concepts are very interesting to examine. Given the assumption that attitudes guide our behavior and taking into account that people can have two attitudes towards the same concept, the examination of people’s attitudes towards a specific nationality could hold benefits for various fields (e.g., politics, academic internationalization). It would enable us to be much more responsive and probably even more persuasive if needed. Imagine, of course totally cliché-ridden, an Italian being early, or a funny German. By confronting people with insight to their attitudes or even acting out the opposite of what they had expected according to priory built up attitudes, one is able to reach levels of intercommunication far beyond the usual. Especially in the case of students’ academic mobility (spending a part of or completing a whole study abroad), the knowledge gathered from examining attitudes towards a specific nation could improve the process of academic internationalization in the long-term.

Since the treaties of Maastricht in 1992, the facilitation of students’ mobility and the teaching of international students is an official part of the European educational policy (West & Barham, 2009). However, albeit a lot of quantitative research is done on student mobility, only few indications exist about levels of satisfaction regarding the time spent abroad. Evidence from the rare qualitative research in this field supports the conclusion that an experience abroad often does not live up to student’s expectations (Olivas & Li, 2006). Among other reasons, this discrepancy between the goals of exchange programs like Erasmus and reality goes



back to extensive problems with cultural differences and to some extent how the students from abroad are welcomed. This provides a clear link to attitudes, since we know that they can guide our behavior. Especially in the case of German students who tend to have a significantly higher academic mobility than students from other comparable countries (Isserstedt & Schnitzer, 2005) it is in particular interesting to examine the single target “nationality” among host countries.

The present study therefore explored the distribution of dual attitudes among three different nationalities with a special emphasis on the employment of a single-target category. The construction of the hypotheses is based on literature in this field. Important insights of past attitude research which warrant the present study can be summarized as follows: a) The conceptual composition of attitudinal dissociation remains unclear. Further research employing paradigms that address the methodological weakness of measuring especially implicit attitudes only in comparison with contrasting categories is needed; b) A scientific interest in concepts of social cognition that are hard to examine (e.g., nationality concepts) is justified, given the lack of qualitative research (e.g., in the field of student mobility).

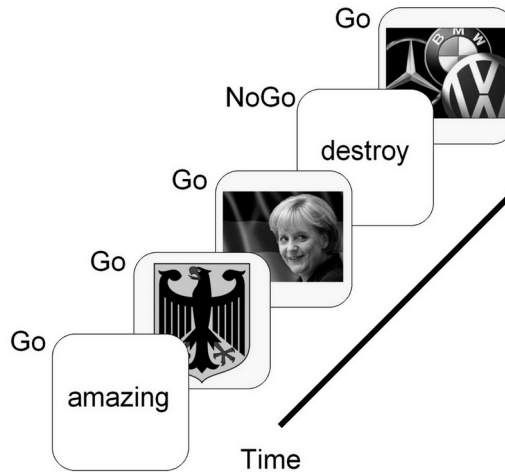
This study aimed to contribute behavioral evidence to the debate of attitudinal dissociation. Among others, a possible solution to this issue is provided by the Go/No Go Association Task (GNAT) (Nosek & Banaji, 2001). The paradigm alleviates the aforementioned weakness of the IAT by the option to implement a single target category without a contrasting category. It has been shown to provide equally reliable outcomes regarding the measurement of implicit attitudes (Nosek & Banaji, 2001). Moreover, with the employment of the GNAT with a single target category, the present study addressed the gap of qualitative research in the course of student mobility by means of assessing attitudes towards a specific nationality. The concept “German” was examined in an Inter-European context of growing host countries for German students going abroad: the Netherlands and Poland (Isserstedt & Schnitzer, 2005). By assessing the implicit and explicit attitudes of Polish, Dutch, and German students, indications for a confirmation or confutation of the dual-representation interpretation could be gathered. Moreover, valuable information about the attitudes towards the concept “German” among Polish and Dutch students was obtained. It was expected that (1) the observed tendencies in implicit and explicit attitudes towards the target-concept “German” differ in their valence. (2) An observable difference between the averages of implicit and explicit attitude assessment among the three sample groups exists.

## METHOD

The present study used a between-subjects design with as independent factor ‘nationality’ (Dutch, Polish, German), and two separate dependent measurements of the implicit and explicit attitudes respectively towards the concept “German”. 61 participants from Poland, the Netherlands, and Germany completed the study. However, four cases from the Polish sample had to be excluded due to invalid trials caused by apparent language comprehension difficulties. Therefore, behavioral

data of 57 participants (mean age 22. 7; 38 female, 19 male) was further analyzed. The age and gender distribution of the individual samples appears as the following: the Netherlands: 18 participants, mean age 23, 11 (11 female, 7 male), Poland: 19 participants, mean age 22, 47 (9 female, 10 male), and Germany: 20 participants, mean age 22, 45 (18 female, 2 male).

Implicit attitudes were explored by means of averaged response latencies on a computer based Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001). The GNAT assesses the strength of mental association between a target concept and two evaluative concepts by measuring the latency of responses towards pairings of stimuli. That is to say, the participant is asked to respond to a certain type of pairing per condition (e.g., when the stimuli displayed is either German or good) while not responding to other pairings (e.g., German and bad). It is assumed that faster responses are directly linked to stronger mental associations.



**Figure 1** - The Go/No-Go Association Task. Participants are provided with stimuli representing the target- category German or a word (positively or negatively valenced). Depending on the condition assigned (here German and good), the participant is asked to press the space bar if the stimuli displayed is either German or good while not responding to negatively valenced words or generic distracter items.

Explicit attitudes were examined with the aid of averaged corrected responses on a Feeling Thermometer, which can be considered as a standardized tool in attitude research (Wilcox, Sigelman, & Cook, 1989). Participants were asked to indicate the “warmness” of their feelings towards the concept “German”. Thereby, the scale of temperature indication ranged from 0° (“very cold”) to 100° (“very warm”) whereas 50° indicated no feeling at all towards the social object in question (e.g., Oskamp & Schultz 2005).

The study was approved by the ethical committee of the Faculty of Psychology and Neuroscience of Maastricht University and abides by the ethical guidelines for psychological research handled by the Faculty of Psychology, Warsaw University.

## Procedure

After a brief oral introduction about the study, the participants had the possibility to ask questions and were informed about their right to stop at any time. In the following, the participants were asked to carefully read through the written instructions of the GNAT on a computer screen. After a practice trial, the participants were left alone to complete four experimental trials within two conditions: congruent trials reflecting the expected attitude towards the concept “German” (pairing of German + bad) and incongruent trials (pairing of German + good). Every trial consisted of 100 randomly presented stimuli on which the participant had to respond by responding to the congruent pairings (Go – responses) while withholding their responses to incongruent pairings (No-Go – responses). In this case each stimuli display automatically ended after 1500ms. Between each trial a short break was provided. At the end of trial four, a written instruction was displayed, signalling that this part of the experiment was finished. The participants were instructed to fill in a demographic survey and to complete the Feeling Thermometer reporting the warmth of feeling toward the concept “German” and their own nationality. Finally, the participants were debriefed and dismissed. If entitled and demanded, the participants were rewarded with 0.5 course credit. The approximate duration of the whole experiment was about 30 minutes.

## Statistical Analysis

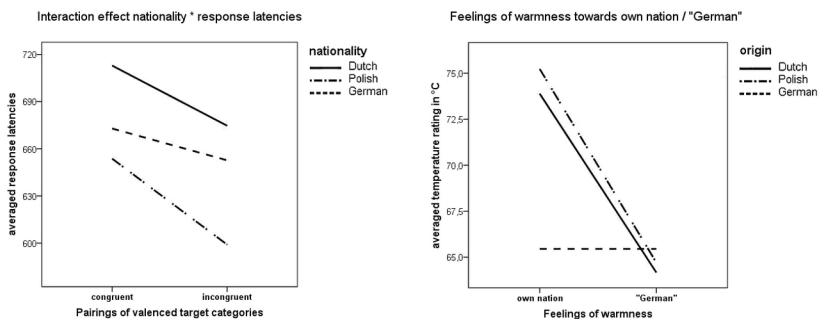
The behavioral data obtained from the measures of implicit (GNAT) and explicit (Feeling Thermometer) attitude towards the concept “German” was prepared for further statistical analysis. In the case of the GNAT, the obtained response latencies were recomputed by means of trial validity and normality of response. Only valid responses (i.e., correct Go- responses to target items), not exceeding the mean  $\pm 2$  *standard deviations* were taken into account. Subsequently, the four experimental trials were recomputed into two independent levels of implicit evaluation (congruent (German + bad), incongruent (German + good)). For the explicit attitudes, the measurements of feeling of warmth towards the concept “German” and “Own nation” were merged into two separate indices of explicit attitude.

A GLM repeated measures was applied to the behavioral data obtained from the GNAT with “nationality” serving as independent between-subject factor and the two levels of implicit evaluation as independent within-subject factors. The averaged response latencies in the congruent and incongruent condition were taken as dependent within-subjects factors. For the analysis of the assessed data from the Feeling Thermometer, again a GLM repeated measures was applied. “Nationality” served as independent between-subject factor whereas the dependent within-subject factors were the averaged reported feelings of warmth towards the concept “German” and the averaged reported warmth of feelings towards the participants’ own nation. An *alpha* level of 0.05 was used for all inferential analyses. In the case of significant main effects, post-hoc comparisons with Bonferroni correction were applied.

## RESULTS

A significant interaction effect of nationality ( $F(2, 1) = 5, 62, p = .004$ ) on the evaluation towards the concept “German” as well as a significant main effect of nationality ( $F(2, 1) = 76, 91, p < .000$ ) was found. In contrast, no significant effect of nationality was found on the explicit attitudes scale (Figure 2). Interestingly, all participants, reacted significantly ( $p < .000$ ) faster in the congruent (German + bad) condition than in the incongruent (German + good) condition. However, among groups, clear differences were observable with the Polish sample reacting the fastest to the congruent target categories ( $M = 599\text{ms}$ ), followed by the German sample ( $M = 653\text{ms}$ ), and the Dutch sample reacting the slowest ( $M = 675\text{ms}$ ). Thus, the response pattern indicates that a negative automatic attitude towards Germans is clearly influenced by the origin of the participant. This seems to especially account for the case of the Polish sample, since the mean difference between the Polish and the Dutch as well as the German sample was more significant than the differences between the Dutch and the German sample (Dutch/German:  $p = .038$ ; Polish/Dutch/German:  $p < .000$ ) as analyzed in the post-hoc comparisons.

An observable difference between the averages of implicit and explicit attitude assessment among the three nationalities was found for implicit, but not for explicit attitudes ( $F(2, 1) = 2, 11, p = .131$ ). However, an obvious trend is visible (Figure 2). As it could have been expected in advance, Germans reported the highest temperatures towards themselves. The least warm feelings, however, were reported by the Dutch sample which was in fact the slowest on the automatic evaluation of the negatively valenced target category. Moreover, a significant main effect of “feeling” could be observed ( $F(2, 1) = 8, 07, p = .006$ ). That is to say, the temperature rating of the Polish and Dutch sample regarding their own nation turned out to be significantly warmer than the temperature rating of the German sample.



**Figure 2** - Interaction effect between nationality and averaged response latency on two independent within-subject factors. 1 = incongruent, 2 = congruent (left); Mean Plots depicting averaged temperature ratings for “own nation” = 1 and “German” = 2 among the three groups compared (right).

## DISCUSSION

The purpose of the present study was to examine implicit and explicit attitudes towards the concept “German” exemplarily for single target evaluations in implicit social cognition. Moreover, the study was set up in a way that it attempted to provide possible solutions to methodological issues that have been identified in earlier studies. The study aimed to broaden the explanatory power of attitude research.

It was expected that the tendencies observed in the assessment of implicit and explicit attitudes towards the concept “German” differ in their valence among three groups with different nationality. Overall, this expectation was confirmed by means of a significant main effect of nationality on response latencies in the assessment of implicit attitude. All participants reacted significantly faster on the congruent pairings (German + bad), being indicative for a negative implicit evaluation of the target concept. However, among groups, significant differences in response latencies were visible. This suggests a broader distribution of implicit attitudes within the overall negative evaluation. The Polish sample group displayed a more negative implicit evaluation of the concept “German” than the German sample. The Dutch sample showed the least negative but still negative implicit evaluation.

In contrast, no significant effect of nationality was found on the explicit attitudes scale, the averaged warmth of feeling towards the concept “German”. All groups unanimously stated a certain, not significantly differing, feeling of warmth towards the concept “German” which is indicative for a neutral to slightly warm explicit attitude (means are grouped around 65°C). Thus, the clear difference that could be observed between the implicit and explicit attitude confirm the expectation of differently valenced attitudes among the three groups.

Moreover, it was expected that an observable difference between the averages of implicit and explicit attitude assessment among the three nationalities exists. The current study fails to provide such results for explicit, but not for implicit attitudes towards the concept “German”. Even though an obvious trend of different warmth towards their own nation and the concept “German” is visible (Figure 2), the outcomes are not differing significantly.

However, even though partially failing to provide support for the second expectation, this outcome is interesting, since a non-significant outcome at this point does strictly speaking provide additional support for the first expectation. That is to say, if all groups evaluate the target concept “German” approximately the same on one scale, but not at all on the other, it is safe to assume that their valence differs to great extent.

Taken together, a distinction of underlying attitude constructs fits the data better than explanations that point to one single construct that has only been assessed with different methods. Thus, a strong indicator for a dual viewpoint of underlying attitude constructs as suggested by a dual model of attitudes (e.g. Wilson et al., 2000; Strack & Deutsch, 2004). However, this might only be true on first glance. Before drawing a definite conclusion with regard to the data obtained, a critical review of additional implications has to be made. Indeed, the response patterns fit a dual attitude model better than a unimodal model. Nevertheless, no evidence was gathered that it is not a unimodal model. Just by gathering indications

for a dual representation model of attitudes, it is by implication impossible to rule out a single-representational model. In line with the elegant re-evaluation of numerous studies within the field of attitudes by Nosek and Smyth (2007) who claimed that irrespective their source implicit and explicit attitudes have essentially independent features and none of the two possible explanations in question does fully account for the abundant depiction of the findings, the present study can only contribute another indication, not a valid proof for a dual model. However, all studies re-evaluated by Nosek and Smyth (2007) suffered from the numerous methodological issues that have been discussed before. Therefore, the observed tendencies in the present study might be considered as having a bigger explanatory impact than previous studies since it particularly addressed multiple of those issues at the same time.

Even though this study aimed to rule out certain methodological errors, one still has to take additional indications into consideration. Greenwald and Nosek (2009), for instance, ascertained that no empirical evidence at all can be used to explain the phenomenon of attitudinal dissociation in terms of underlying structural disparity. Due to the fact that attitudes are still a subject to theoretical approaches, they remain “hypothetical and unobservable” (Greenwald & Nosek, 2009, p.9) and can therefore not be addressed with the aid of behavioral research. However, empirical evidence is still used to examine attitudes on a regular basis. This dissension about the usefulness of behavioral evidence in attitude research fuels the controversy about the attitudinal dissociation phenomenon only further. Therefore, it is pivotal to use Nosek and Smyth’s (2007) theoretical, but at least fully accountable explanation of this phenomenon together with behavioral data, the H<sub>2</sub>O example: Attitudes might be like the observably different forms of water, like ice, snow, and steam, but deriving from a single molecular structure. The different forms of attitudes might be triggered, just as in water, by situational factors as temperature. However, even though the underlying molecular structure (or the underlying attitudinal concept) might be the same, no one would ever try to swim in a snowdrift. The same goes for attitudes. They might or might not derive from the same structure, we do not know that. However, in real life, it is much more suitable to study them as structurally different. Taking this implication into account, it could be even the case that every observed tendency in favor of a dual viewpoint is a misconception. However, as mentioned before, a unimodal theoretical model can by implication not rule out a dual viewpoint. Therefore, even though providing us with a satisfactory explanation, the ongoing controversy will not be solved by this statement.

Next to the rather theoretical accounts that could weaken the conclusions drawn from the present study, this research faces several practical limitations. First of all, even though the stimuli used in the Go/No-Go Association Task (GNAT), were carefully reviewed, a general lack of sufficient comprehension of English was observed in the Polish sample. Obvious cases of misinterpreting text stimuli were excluded. However, it could still be the case that the level of comprehension influenced the responses. In future research, this issue should be accounted by assessing in the native language of participants only.

Second, the GNAT paradigm suffers from the methodological problem that



it accounts for measuring single attitude objects, but there is no possibility of proving that the attitude object was evaluated independently from other factors. Therefore, it could be the case that the results are biased in terms of e.g., contextual factors. This limitation is in line with the review of Blair (2002) who claims that “automatic attitudes - like self-reported attitudes - are sensitive to personal, social, and situational pressures (...)” and “do not provide a ready solution to the problem of attitude malleability” (p. 256).

Third, some additional methodological issues that should be taken into consideration in future research are connected to the paradigms chosen. The GNAT only accounts for stimulus categorizations. Different facets of an attitude object were shown and had to be categorized into the target category. In turn, the assessment of explicit attitudes with the Feeling Thermometer is a category label evaluation. This could account for the observed tendency contrarities of those two measurements equally good as really different underlying attitude constructs. Furthermore, as mentioned above, the Feeling Thermometer is a solely category label evaluation that depicts direction and intensity on a continuous scale. By doing so, individual differences producing variance have to be taken into account. According to Wilcox, Sigelman, & Cook (1989), some people tend to display overall “warmer” reports than others and some tend to use broader partitions of the scale. Next to that, a certain amount of unexplained variance is always expected, which cannot be attributed to people’s tendencies but rather to random noise and other factors (e.g., contextual factors and/or biases). However, this limitation loses impact given that research has shown that the foremost predictor of the Feeling Thermometer is indeed the attachment to the concept questioned (Alwin, 1997). Eventually, the calculated power criterion was not met. The samples used are very small and only 57 participants could be included into the analysis. Given the strong significance of the results, however, this limitation can be considered as minor.

In spite of these limitations, the current study provides valuable information about the employment of a single-target category, here nationality. The controversy stated above might not be solved and given the prudential analysis provided by Greenwald and Nosek (2009) probably will not be resolved in the near future. However, despite this controversy, on one topic unity remains. Attitudes - also implicit - are under certain circumstances important explanative and predictive factors of people’s current and future behavior (Bohner & Wänke, 2002). This conjuncture allows adding further relevance: The findings of the present study can be used in the additional focus of this paper outlining practical implications for the process of internationalization on university level.

The Netherlands as well as Poland are growing host countries for German mobile students. Almost half of the students starting a Bachelors’ program at Maastricht University in 2010 were coming from abroad. Furthermore, there is indication for an ascertained success on recruiting German students within that population (Maastricht University, 2010). Similar patterns indicate a host country tendency for Poland. For instance, at the University of Warsaw, Germans represent the highest number of incoming student’s for short term studies (International Relations Office, n.d). Given these figures, it is futile to assume that deeply ingrained implicit negative evaluations towards the concept “German” appearing as strong



as shown in this study, would not be influential on studying together. Therefore, one should clearly consider the development of internationalization concepts that deepen the often shallow, image-conscious approaches handled by universities nowadays. These approaches, including the care for international students' needs in terms of housing, lowered language barriers, and organized get-togethers, self-evidently for international students only, usually even facilitate the ostracism of incoming international students. Naturally, this is not what universities sought to accomplish, however, a more sustainable approach of international student integration would be certainly more beneficial. Universities should work hand in hand with all students, local as well as international to establish concepts of internationalization that provide a real cross-cultural cooperation. Possible short-term objectives could for instance cover the enlargement of teambuilding activities and/or the extension of mentor programs, pairing up local and incoming students for longer periods of time than the common introduction week. Such provisions might enable students to override cultural barriers and even could lead to attitude changes that contribute substantially to an improvement of the experience abroad.

Whereas the practical application of the findings in this paper provides us with relatively easy to achieve short-term objectives that could actually make a difference, the theoretical implications remain limited. The controversy, on the origin of our implicit and explicit attitudes clearly demands further research. This study conducted in the naturalistic environment of academic student mobility has only begun to shed more light on the attitudinal dissociation phenomenon by providing some evidence in favor of a dual viewpoint. The additional, practical focus of the present study, concerning the process of internationalization, however, adds a perpendicular relevance in terms of potentials and pitfalls of academic internationalization allowing the fragmentary, but meaningful conclusion that much like students, regardless their precise origin, our attitudes towards other people are meaningful and influential.

### **Acknowledgement**

“Qui dedit beneficium, taceat. Narret, qui accepit.” —Seneca

Many thanks to Dr. Arie van der Lugt for his excellent supervision and to Frauke A. Holstein for her essential advice at all stages of this project.

### **REFERENCES**

- Alwin, D. F. (1997). Feeling thermometers versus 7-point scales. *Sociological Methods & Research*, 25, 318-340.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242-261.
- Bohner, G., & Wänke, M. (2002). *Attitudes and Attitude Change*. New York, NY: Psychology Press.
- Bolten, J. (2006). Die Entwicklung von Nationalstereotypen im Globalisierungsprozess. Hypothesen zum Auftakt einer international durchgeführten Langzeituntersuchung zu

- Veränderungen des Deutschlandbildes bei Studierenden. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 11, 4-21.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual review of psychology*, 54, 297-327.
- Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, 40, 535-542.
- Greenwald, A. G., & Nosek, B. A. (2008). Attitudinal dissociation: What does it mean? In Petty, R. E., Fazio, R. H., & Briñol, P. (Eds.), *Attitudes: Insights from the new implicit measures* (p. 65-82). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes & Intergroup Relations*, 11, 69-87.
- International Relations Office, University of Warsaw. (n.d.). *University of Warsaw - Erasmus in 1998 – 2011. Incoming Short-term Students according to countries*. Warsaw, Poland: International Relations Office, University of Warsaw.
- Isserstedt, W., & Schnitzer, K. (2005). Internationalization of higher education: foreign students in Germany, German students abroad. *Berlin: Federal Ministry of Education and Research*.
- Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preferences and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin*, 30, 629- 642.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, 81, 774.
- Maastricht University. (2011). *Annual Report 2010*. Maastricht, the Netherlands: Maastricht University.
- Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, 19, 395-417.
- Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition*, 19, 625-666.
- Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, 54, 14-29.
- Olivas, M., & Li, C. (2006). Understanding stressors of international students in higher education: What college counselors and personnel need to know. *Journal of Instructional Psychology*, 33, 217.
- Oskamp, S., & Schultz, P. W. (2005). *Attitudes and opinions*: Lawrence Erlbaum.
- Steffens, M. C. (2005). Implicit and explicit attitudes towards lesbians and gay men. *Journal of Homosexuality*, 49, 39-66.
- Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, 8, 220-247.
- Wehmeier, S., & Ashby, M. (Eds.). (2000). *Oxford Dictionary of Current English*. Oxford: Oxford University Press.
- Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social issues*, 25, 41-78.
- Wilcox, C., Sigelman, L., & Cook, E. (1989). Some like it hot: Individual differences in responses to group feeling thermometers. *Public Opinion Quarterly*, 53, 246-257.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-126.

# To what extent do the various measures of confidence affect the accuracy-confidence relationship in earwitnesses? A review of research on the earwitness testimony

## REVIEW

Despite inconsistent findings regarding a relation between accuracy and confidence, the use of confidence as a correlate of accuracy is a common subject in forensic psychology research. Therefore, in this review the existence of an accuracy-confidence relationship (A-C relationship) in earwitnesses is examined by considering several variables influencing earwitness memory. First, two different methods of assessing confidence are discussed, the discrimination method and the absolute accuracy method, indicating no differences between these two kinds of methods. Subsequently, confidence is discussed in relation to incorrect/correct decisions and target present/absent conditions. It appears that the mere behavioural act of making an identification (vs. rejecting identification) increases the individual's perceived confidence, regardless whether this identification is correct or not. Finally, future directions for research based on the findings of this review are highlighted.

**Keywords:** literature study, earwitnesses, accuracy-confidence relationship, line-ups

Annette Verhaeg; Master student Neuropsychology  
Maastricht University, Maastricht, the Netherlands

---

[a.verhaeg@student.maastrichtuniversity.nl](mailto:a.verhaeg@student.maastrichtuniversity.nl)

## INTRODUCTION

When a crime occurs, it is common that an eyewitness is asked to attempt to identify the perpetrator of the crime in a line-up. However, sometimes there are no eyewitnesses who have witnessed the crime, but only earwitnesses. In these situations, the only information available about a crime is the memory of hearing one or more voices. Although it might be thought that people will never forget the voice of a perpetrator, evidence shows that the recognition of a voice can be very difficult (Read & Craik, 1995; Van Wallendael, Surace, Parsons, & Brown, 1994) and that earwitnesses are less accurate in identifications than eyewitnesses (McAllister, Dale, & Keay, 1993). Why is voice identification that difficult?

Voice identification is difficult because of the many factors that influence our memory of a voice. These factors can be divided into three different variable categories, speaker variables, procedural and situational variables, and listener variables (Yarmey, 1995). The influence of these variables is investigated in a lot of studies, but it remains difficult to study only one variable in particular because one variable is almost never independent from other variables. To get an idea of the different variables within the three categories, some examples will be discussed with regard to the categories.

The first category, speaker variables, consists of differences in the voice of the perpetrator between crime and line-up. This category incorporates variables like voice distinctiveness (Mullennix et al., 2011; Saslove & Yarmey, 1980), voice familiarity (Read & Craik, 1995; Yarmey, 1995), accent (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2006) and voice disguise by whispering (Orchard & Yarmey, 1995), which are all found to influence voice recognition.

The second category, procedural and situational variables, includes variables that differ between interrogation situations. One of these variables is the length of the interval between the witnessing itself and voice identification. Many studies compared the effects of different retention intervals, but evidence in this topic is not unambiguous (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004; Kerstholt, et al., 2006; Saslove & Yarmey, 1980; Van Wallendael, et al., 1994). Therefore, it is very difficult to define the optimal retention interval. Other variables within the procedural category are about the effects of being both an earwitness and an eyewitness (McAllister, Dale, Bregman, McCabe, & Cotton, 1993; Stevenage, Howland, & Tippelt, 2011), and about the effects of the line-up, in which topics like one-person versus many-person line-ups (Yarmey, 1995), number of voices in the line-up (Bull & Clifford, 1984) and the effects of feedback (Quinlivan et al., 2009) are emphasized.

The third category, listener variables, is about the witness who may feel different in two situations. This could for instance refer to effects of stress and arousal on voice recognition evoked by the presence of a weapon (Yarmey, 1995). However, a disadvantage in studies investigating these effects is the ecological validity. For ethical reasons participants in such studies cannot be a direct witness of the crime, but just watch a crime on a TV screen, which is not comparable to being witness of a real crime. This could influence the degree of stress and arousal. The listener variable category furthermore includes the effect of confidence (or certainty). "It

seems intuitively plausible that a person is more likely to be correct when he or she is certain of being correct” (Bull & Clifford, 1984, p.104). This assumption is, however, not invariably confirmed, as Bull & Clifford also indicate. There are studies which have found a relationship between the accuracy of an identification (correct or incorrect) and the confidence of a person on his or her identification (Saslove & Yarmey, 1980; Van Wallendael, et al., 1994). However, on the other hand, some studies are published in which no or only a weak relationship is found (Kerstholt, et al., 2004, 2006; Yarmey, 1986).

Despite inconsistent findings regarding a relation between accuracy and confidence, and a warning to be cautious to infer accuracy from confidence, the use of confidence as a correlate of accuracy is a common subject in forensic psychology research (Boydell & Read, 2011). This emphasizes why it is still important to evaluate the existence of an accuracy-confidence (A-C) relationship. Many factors, such as the variables described before, could have an influence on the A-C relationship, but also the measurement of confidence itself could affect this relation (Van Wallendael, et al., 1994). Therefore, this article will discuss some findings about the relationship between accuracy and confidence in detail. The central research question in this article will therefore be: To what extent do the various measures of confidence affect the accuracy-confidence relationship in earwitnesses? Different measurements of confidence will be discussed with respect to the following issues:

1. Does the method of measurement influence results?
2. Correctness and line-up: Do these variables explain differences in confidence rates? In this section, two main issues are discussed:
  - a) Does a difference in confidence exist between participants who make an accurate and participants who make an inaccurate identification?
  - b) Does a difference in confidence exist between participants who are in a target present condition and participants who are in a target absent condition (line-ups in which the suspect is respectively presented or not)?

The answers to these questions will help in giving a more concrete insight to the strength of the A-C relationship in earwitness identifications. This could also help jurors to make better decisions about a witness with a higher or lower degree of certainty.

## DOES THE METHOD OF MEASUREMENT INFLUENCE RESULTS?

Several ways of measuring confidence are used in the studies testing the A-C relationship in earwitnesses. In some experiments, participants first have to listen to a line-up consisting of several voices. Then they have to decide if a target voice (which was heard during an earwitness situation) is present in this line-up. After this decision participants have to rate how confident they are about their choice (e.g. Orchard & Yarmey, 1995). However, in other experiments subjects are told that they will listen to a line-up in which they will hear one voice at a time. In this case, subjects have to record their judgment after each voice as to whether the person is the perpetrator or not, and how confident they are of their decision (e.g. Saslove

& Yarmey, 1980). Van Wallendaël and his colleagues (1994) explain that the two methods lead to two different kinds of information about confidence. When the first method is used, the absolute accuracy is considered. "This is defined as the subject's final choice of which line-up voice was the target" (Van Wallendaël, et al., 1994, p. 665). In this way, the participant chooses only one voice after hearing the entire line-up and only rates his/her confidence about this voice. This confidence score will only give information about the voice which is, at that time, mostly experienced as the target voice, and gives no information about how other voices are perceived.

When participants have to judge about each voice in particular, a second method is used: discrimination. Discrimination is a ratio of the rated confidence in all the voices. Thus, someone who gives the target a confidence rate of '5', and all other voices '0' will obtain a perfect 1.00 discrimination score. Given that a discrimination score less than 1.00 might indicate that the subject is not entirely sure about his/her decision, the experimenter might receive more information about the memory and cognitive processes of the participant (Van Wallendaël, et al., 1994). Moreover, the data obtained for each voice component (target/ distractor) allows further analyses of any accuracy changes (Van Wallendaël, et al., 1994). Therefore the use of this method could be of more theoretical interest.

To examine whether these two methods lead to differences in results regarding the A-C relationship, several studies will be discussed in which these different methods are used. Obviously, only the studies which reported results about an A-C relationship are included in this discussion. With regard to the studies that used the absolute accuracy method, only five studies were found (Kerstholt, et al., 2004, 2006; Orchard & Yarmey, 1995; Philippon, Cherryman, Bull, & Vrij, 2007; Read & Craik, 1995). Only four studies were found which used the discrimination method (Saslove & Yarmey, 1980; Stevenage, et al., 2011; Van Wallendaël, et al., 1994; Yarmey, 1986). Differences and similarities between studies are discussed.

### **The absolute accuracy method**

In an experiment by Orchard and Yarmey (1995), people had to listen to six voices in a voice line-up. Afterwards, they had to decide whether or not the perpetrator was in the line-up by selecting the number of the voice, by saying that he was not in it or by saying that they did not know. The participants who chose a voice had to indicate their confidence about their choice on a 5-point scale. The relationship between accuracy and confidence was measured as follows: They used a 10-point accuracy-confidence index. In this method, correct identifications (hits) in target present line-ups and correct rejections in target absent line-ups with a confidence rating of 1, 2, 3, 4, or 5 are scored as 6, 7, 8, 9, or 10, respectively. Incorrect identifications with a confidence rating of 1, 2, 3, 4, or 5 are scored as 5, 4, 3, 2, or 1, respectively. This means that the higher the score, the better the accuracy and confidence in the voice-selection. Altogether, results of all measurements revealed significant A-C relationships (point biserial correlations) for both target present line-ups ( $r = .25$ ,  $p < .001$ ) and target absent line-ups ( $r = .36$ ,  $p < .001$ ). This is, however, the only study which used the absolute accuracy method and which found significant A-C relationships.

Four other studies found no significant A-C relationship (Kerstholt, et al.,



2004, 2006; Philippon, et al., 2007; Read & Craik, 1995). In the research conducted by Read and Craik (1995), participants listened to a 6-person line-up, and had to indicate their certainty on a 4-point scale. Three experiments were provided, but overall, no significant A-C relationship was found. They only found a significant relationship when the voice in the line-up was identical (exact same recording) to the voice in the witness situation ( $r = .25$ ,  $p < .05$  and  $r = .40$ ,  $p < .001$ ). When the line-up voice was rerecorded there was only a slight tendency for an A-C relationship ( $r = .17$ , no p-value reported). However, in reality the voice line-up will never be an exact copy of the voice during the crime. Moreover, this tendency for a relationship was only due to an increase in certainty for correct choices, but not for incorrect choices. Therefore they concluded that the A-C relationship was not strong enough to trust on it.

Two very similar experiments by Kerstholt et al. (2004) and by Kerstholt et al. (2006) again found no significant A-C relationship. In both studies people listened to 6 voices and indicated their confidence ratings on a 7-point scale after hearing all the voices. In order to investigate whether the accuracy of the judgements could be predicted by the confidence judgement of the participant, a logistic regression analysis was carried out. In both experiments no relationship was found. However, it is notable that in both articles no r-values and p-values are reported. Finally, a research conducted by Philippon et al. (2007), found no significant relationship ( $r = 0.094$ ,  $p = .475$ ). In this study point biserial correlations were used to measure the A-C relationship. Participants listened to a 6-person line-up, and rated their confidence on a 5-point scale.

In conclusion, most studies using the absolute accuracy method found no strong A-C relationships and overall this relationship was found by assessing the point biserial correlation. However, it is very difficult to compare the results of all these studies, because very different scales are used to assess the person's confidence, ranging from 4- to 7-point scales. This could lead to differences between results, as people differ in their answers depending on the scale length (Forzano & Gravetter, 2009). Moreover, it is remarkable that only the study of Orchard and Yarmey (1995) transformed the raw confidence rates in ten new scores by using an accuracy-confidence index. In this way, the fact of making an accurate or inaccurate identification was taken into account when calculating the A-C relationship. The biserial correlation calculated by using these scores is different from using the raw scores. With regard to the four other studies, it is clear that they also used a biserial correlation, but it is not clear if they used such index as well. This could produce differences in the final A-C correlation. These issues make it difficult to draw a definite conclusion about whether the use of the absolute accuracy method leads to a significant or insignificant A-C relationship.

### **The discrimination method**

In a research conducted by Saslove and Yarmey (1980) participants listened to a 5-person line-up and judged whether each voice was old (perpetrator) or new (innocent person). In addition, subjects had to indicate their confidence level in their old/new decisions on a 3-point scale (possible, probable, or certain). For each participant the line-up consisted of one target voice and four distractor voices.



Recognition memory was analysed in terms of hit - miss scores at one scale and false alarms - correct rejection scores at a second scale. The hit - miss scale consisted of six possible scores ranging from certain miss (score 1) to certain hit (score 6). Therefore, the maximum score on this scale was six. On the other hand, the false alarm - correct rejection scale also consisted of six possible scores ranging from certain false alarm (score 1) to certain correct rejection (score 6). This was evaluated for each of four distractor voices. In this way the maximum score on this scale was 24 (4 times a certain rejection score of 6). The higher the score on each scale, the more accurate is the response to the voices. A point biserial correlation was used to assess the relationship between certainty (total score of a participant) and accuracy (correct identification or not). A small but significant correlation was found ( $r = .26$ ,  $p < .01$ ).

Another study which used the discrimination method was the experiment by Yarmey (1986). Similar to the study by Saslove and Yarmey (1980), this experiment used a five-person line-up. However, participants were not told how many voices they would hear. Again the participants had to indicate whether the voice they heard belonged to the perpetrator or not. Additionally, subjects had to indicate their certainty for each decision on a 4-point scale. Although this research used the discrimination method as well, the point biserial correlation between confidence of response on correct identifications and confidence of response on incorrect identifications was not significant ( $r = -.003$ ,  $p$ -value not reported).

It is notable, however, that Yarmey (1986) did not completely describe which scores of confidence he used in the correlation measurement. It is not clear, whether he also used a total certainty score as a measurement for confidence, as used in Saslove and Yarmey (1980), or whether they used another scoring method. Furthermore, in this experiment the participants did not only hear the perpetrator during the observation of the assault, but also saw the perpetrator. This makes a comparison difficult, since visual information can interfere with auditory information (McAllister, Dale, Bregman, et al., 1993).

In the study by Stevenage, Howland, & Tippelt (2011), participants were randomly exposed to a dual-input (audio and visual) or to a single-input condition (visual only/ audio only). In this way, it was possible to investigate the influence of interference on the A-C relationship, as described above. The results of McAllister, Dale, Bregman, et al. (1993) were replicated, although this did not lead to differences in significance of the A-C relationship. The results showed neither a significant A-C relationship for the dual-input ( $r = .15$ ) nor for the single-input condition ( $r = .13$ ) at the 0.05 significance level. These findings make a comparison between Saslove and Yarmey (1980) and Yarmey (1986) more meaningful.

Three striking points, with regard to the study by Stevenage, Howland, & Tippelt (2011) are noticed. First, in this research they use mean confidence scores to examine the A-C relationships. However, as in the experiment by Yarmey (1986), it is not completely clear how they calculated these mean confidence rates. Secondly, they used a 7-point scale to assess the confidence of participants, which is very different from the 3-point scale used by Saslove & Yarmey (1980) or the 4-point scale used by Yarmey (1986). The remaining point to notice is about the voices that had to be remembered. In this study participants had to remember eight studied voices instead of one perpetrator's voice. As indicated in the study, this could also

influence the results, because the task of studying and later identifying eight targets it is very different from identifying a voice which is incidentally learned.

Finally, Van Wallendaël et al., (1994) also discussed the discrimination method. In this study the A-C relationship of earwitnesses was measured in three ways. Before listening to the line-up, subjects had to indicate their confidence (on a 7-point scale) in their ability to recognise the voice. They then listened to the line-up, during which they rated each voice on a scale from 0 (sure, this is not the voice) to 6 (sure this is the voice). After listening to all the voices, participants were asked to choose the voice that they believed to be of the perpetrator. In addition, they had to indicate if they would swear to this identification in a court of law. The results revealed no significant relationship between subjects' pre-line-up confidence and the actual performance in the recognition task. This was measured by comparing the average pre-line-up confidence ratings of accurate and inaccurate participants. The relationship between accuracy and post-line-up confidence in the chosen voice was significant ( $p < .01$ ), in that accurate subjects showed higher confidence ratings than inaccurate subjects. Furthermore, a greater accuracy was found for subjects who were willing to swear to their identification ( $p < .001$  and  $p < .05$  for respectively target present and target absent line-ups). Unfortunately, it is not completely clear how the mean confidence scores were estimated. The researchers asked people to indicate their confidence level on each voice, but it seems that they only used the confidence score of the chosen voice in their assessment of the A-C relationship. If this is the case, they did not use the discrimination method, but the absolute accuracy method.

In conclusion, two of four discussed studies seemed to find correlations between accuracy and confidence by using the discrimination method. An explanation for these ambiguous findings might be sought in the way in which these studies use the discrimination method. Not every experiment exactly clarified how they computed the A-C relationship. It is often not clear how they used all confidence scores (on each voice) in their assessment of an A-C relationship. Many studies report a 'mean' confidence (Yarmey, 1986), but do not explain how this mean confidence rate is calculated. Furthermore, remarkable differences are found in the length of the confidence scales used to measure the certainty of participants, the number of voices in the line-up, and the use of one voice versus more voices that have to be remembered. Due to these differences, no definite conclusion can be made about whether the use of the discrimination method mostly leads to a significant or insignificant A-C relationship.

### **Conclusion absolute accuracy and discrimination methods**

In summary, it has become clear that there are considerable discrepancies between the discussed studies within each method. This makes it very difficult to conclude whether the discrimination method and absolute accuracy method really differ in results about an A-C relationship. To investigate whether these methods really differ from one another, future research has to compare more similar experiments. Furthermore, calculations, with regard to the confidence rates used to assess the A-C relationship, must be more clearly defined. If it turns out to be that those two

methods do not differ, it might be better to use the discrimination method, since this method contains rates of confidence for every voice separately. This allows further analyses and could give us more insight into the memory processes of the participants (Van Wallendael, et al., 1994).

### CORRECTNESS AND LINE-UP: DO THESE VARIABLES EXPLAIN DIFFERENCES IN CONFIDENCE RATES?

Most studies and experiments measuring the A-C relationship use mean confidence rates of all participants in their calculation (e.g. Saslove & Yarmey, 1980). However, sometimes it is better not to merge all participants in measuring relationships, because this could give a distorted view of real situations. Confounders could cause relationships to be overestimated or underestimated. An overestimation occurs when a correlation seems to exist when all participants of an experiment are generalised, but disappears when some variables (confounders) are controlled for. An underestimation occurs when no correlation seems to exist, but by controlling for some variables, actually some relationship appears. In calculating the A-C relationship in earwitnesses such over- or underestimating could be present, because in most studies all participants (incorrect, correct, participants in target-present and in target absent conditions) are merged in the analysis.

Read and Craik (1995) found that controlling for a variable did have an influence on results by producing different conclusions. In their study people first had to listen to a target voice and then listened to either recordings of conversations (not the same as initially heard), an identical line-up (exactly same recording as initially heard target voice) or a rerecorded line-up (rerecording of the initially heard target voice). After making a voice identification, participants had to indicate their confidence in their decision on a 4-point scale with four representing high confidence. Afterwards, mean confidence rates (C) were calculated in each condition separately for correct and incorrect decisions. The A-C relationship was assessed by calculating whether C of correct and incorrect responses significantly differed from each other. A significant difference between correct and incorrect C rates was found in the identical line-up condition ( $p < 0.001$ ). However, it was remarkable that this A-C relationship only appeared to be significant, because participants had a higher confidence rate for correct identifications in the identical line-up ( $C = 2.63$ ) than when they had to listen to recordings of conversations ( $C = 2.11$ ). The appearance of the significant A-C relationship was not due to lower confidence rates in incorrect identifications in the identical line-up ( $C = 1.88$ ), because those were very comparable to the rates when they had to listen to recordings of conversations ( $C = 1.87$ ). A second experiment, which was conducted to replicate these results, showed the same trend of results (Read & Craik, 1995).

The overall conclusion in the research conducted by Read and Craik (1995) was that the A-C relationship was too small to conclude that it could be used as a reliable instrument to rely on an earwitness or not. However, it can be concluded from the study by Read and Craik (1995) that a small relationship seems to exist when

controlling for the variable, '*correctness*' (incorrect or correct identification). This illustrates why it could be important to take a closer look at the differences within groups, because mean confidence rates do not always show these small differences. Therefore, more studies reporting numbers of confidence rates for correct and incorrect participants separately are discussed in this section.

Furthermore, some studies contain another variable, '*line-up*', in which confidence rates of target present (TP) line-ups and rates of target absent (TA) line-ups are compared. To investigate if these two kinds of line-ups differ in confidence rates, these results are discussed as well.

Only five studies were found (Pickel, French, & Betts, 2003; Read & Craik, 1995; Stevenage, et al., 2011; Van Wallendaal, et al., 1994; Yarmey, 1986), in which numbers about confidence rates were represented separately for '*correctness*' (participants who identified correctly or incorrectly), '*line-up*' (participants who were in a TP line-up, or in a TA line-up), or both. In one of these experiments the participants were both an earwitness and an eyewitness (Pickel, et al., 2003). This study used both a TP line-up and a TA line-up, but reported numbers only for correct (witnesses who correctly identified the target's voice, or who correctly rejected all voices in target absent line-ups) and incorrect participants (witnesses who incorrectly rejected the line-up, or who incorrectly identified a distractor voice). A 7-point scale was used to measure confidence (C) with higher ratings representing more confidence. In the experiment, a marginally significant difference ( $p = .062$ ) was found between correct (mean  $C = 4$ ) and incorrect decisions (mean  $C = 3.54$ ). However, because the participants in this experiment were both eyewitness and earwitness, results could be distorted (McAllister, Dale, Bregman, et al., 1993).

Another study, in which numbers about confidence rates were reported, was a study by Stevenage et al. (2011). In this research, no TA line-up was used, so the results only contained information about the '*correctness*' of the participants in a TP line-up condition. As in the experiment of Pickel, et al. (2003), they used a 7-point scale to measure confidence. Participants' decisions were only rated as correct when the target voice was indicated as being '*old*' (previously heard). Otherwise, their decisions were rated as incorrect. Results revealed a significant difference between correct and incorrect decisions, in which correct decisions showed higher confidence rates than incorrect decisions (respectively,  $C = 4.23$  and  $3.99$ ,  $p = < .001$ ). This significant difference is, however, not comparable to the results that were found in the study of Pickel, et al. (2003). This may be caused by the fact that the confidence rates as reported by Pickel are averages of confidence rates in TP and TA line-ups (confidence rates of hits & correct rejections are merged, as are misses and false alarms), although Stevenage's study only reports information about the TP line-up condition.

A comparison of the two studies (Pickel, et al., 2003; Stevenage, et al., 2011) is impossible until it is proven that confidence rates between TA and TP line-ups do not differ. If confidence rates between these two kinds of line-ups differ, the rates as reported by Pickel, et al. (2003), are poor representations of the variable '*correctness*', because these rates are averages of TA and TP line-ups. To investigate if these line-ups differ in confidence rates, two studies will be discussed that reported information about confidence rates of the two variables '*correctness*' and '*line-up*'

(Van Wallendaël, et al., 1994; Yarmey, 1986). In the experiment by van Wallendaël et al. (1994), participants could indicate their confidence on a 0 to 6 scale, instead of the 1 to 7 scale used in Pickel, et al. (2003) and Stevenage, et al. (2011). The difference between correct and incorrect decisions in this study was significant ( $P < .01$ ). Participants who made correct decisions rated their confidence higher ( $C = 5.456$ ) than participants who made incorrect decisions ( $C = 4.891$ ). However, the difference in confidence rates between TP line-ups and TA line-ups was significant as well ( $p < .001$ ): Participants in TP line-ups indicated higher confidence rates in their decisions than participants in TA line-ups ( $C = 5.364$  and  $4.784$ , respectively).

In line with Van Wallendaël, et al. (1994), the experiment by Yarmey (1986) also showed that confidence rates of participants in the TP line-up were higher ( $C = 3.05$ ) than the rates in TA line-ups ( $C = 2.7$ ). Yarmey (1986) used, however, a 4-point confidence scale, instead of a 7-point scale. Furthermore, differences in confidence rates were found between participants that were correct ( $C = 2.65$ ) and those who were incorrect ( $C = 3.1$ ). However, these results were not complemented by p-values and were very much in contrast to the results found by the other studies (Pickel, et al., 2003; Stevenage, et al., 2011; Van Wallendaël, et al., 1994), because in this experiment the confidence rates of correct decisions were lower (instead of higher) than the rates of incorrect decisions. It is not completely clear why these results differ so much from the other studies, but as Yarmey (1986) used a 4-point scale instead of a 7-point confidence scale, the difference could be due to the difference in scales. As noted before, it is important for future research to investigate the consequences of these differences in confidence scales used.

The fact that both Van Wallendaël, et al. (1994) and Yarmey (1986) found differences in confidence rates between TA line-ups and TP line-ups, indicates that the mean confidence rates for the variable 'correctness' could be distorted by the differences induced by the variable 'line-up'. However, the confidence rates for the variable 'line-up' are mean scores as well, because these scores are the mean of correct and incorrect decisions, separately for TA and TP conditions. Therefore, the results could also be distorted the other way round.

To get more information about this/these distortion(s), confidence rates of Yarmey's study will be re-evaluated in this review. This will be accomplished by first evaluating the value of mean confidence rates for incorrect and correct decisions, after which the value of mean confidence rates for TA and TP line-up will be evaluated. The section will be concluded with an evaluation of splitting up incorrect decisions in false alarms and misses. Table 1 schematically shows how in this article the different mean confidence rates will be calculated from the confidence rates of each particular condition - TA line-up and correct decisions, TP line-up and correct decisions, TA line-up and incorrect decisions, and TP line-up and incorrect decisions. Since the study conducted by Yarmey (1986) was the only study that reported all these confidence rates, only this study can and will be re-evaluated.

**Table 1:** A schematic representation of the calculations made to obtain mean confidence rates (CR).

Correctness of decision	Kind of line-up		Mean CR within correctness
	TA	TP	
Correct	A	B	$(A+B)/2$
Incorrect	C	D	$(C+D)/2$
Mean CR within line-up	$(A+C)/2$	$(B+D)/2$	

*Note.* TA = target absent line-up; TP = target present line-up; CR = rate of confidence; A = mean CR for participants who correctly rejected the line-up in the TA condition; B = mean CR for participants who correctly identified the target in the TP condition; C = mean CR for participants who incorrectly identified a distractor voice in the TA condition; D = mean CR for participants who incorrectly identified a distractor voice or missed the target voice in the TP condition;  $(A+B)/2$  = mean CR for participants who made a correct decision;  $(C+D)/2$  = mean CR for participants who made an incorrect decision;  $(A+C)/2$  = mean CR for participants in the TA condition;  $(B+D)/2$  = mean CR for participants in the TP condition.

### The value of mean confidence rates for incorrect and correct decisions

To evaluate whether mean confidence rates of the variable ‘correctness’ are informative, different confidence rates reported by Yarmey’s study are discussed (see table 2). The first thing to notice is the difference of correct decisions between the TA line-up and the TP line-up. People who correctly chose a voice in the line-up, were far more confident about their decision ( $C = 3.65$ ) than people who correctly rejected the voices in the line-up ( $C = 1.65$ ). For this reason, the mean confidence rate of correct decisions ( $C = 2.65$ ) is not a good measurement of overall correct decisions, as it underestimates confidence rates of participants who make correct decisions in the TP line-up condition, and overestimates correct decisions of participants in the TA condition.

**Table 2:** Mean confidence rates separate for participants in each condition (TA correct, TP correct, TA incorrect and TP incorrect).

Correctness of decision	Kind of line-up		Mean CR within correctness
	TA	TP	
Correct			2.65
Correct rejection	1.65		
Hit		3.65	
Incorrect			3.1 <sup>a</sup>
False alarm	3.75	3.5	
Miss		1.4	
Mean CR within line-up	2.7	3.05 <sup>b</sup>	

*Note.* TA = Target absent; TP = Target present; CR = confidence rate. Adapted from “Earwitness speaker identification”, by A.D. Yarmey, 1986, *Psychology, Public Policy, and Law*, 1, 792-816. <sup>a</sup>Calculated by first averaging the CR for ‘incorrect decisions in TP condition’ (mean of false alarm and misses in the TP condition = 2.45) after which the incorrect confidence rates for TA and TP condition are averaged  $[(2.45+3.75)/2]$ . <sup>b</sup>Calculated by first averaging the CR for ‘incorrect decisions in TP condition’ (mean of false alarm and misses in the TP condition = 2.45) after which the TP confidence rates for incorrect and correct decisions are averaged  $[(2.45+3.65)/2]$ .



The same effect is found to be true in the evaluation of the mean confidence rate of incorrect decisions. Therefore, it can be concluded that it is important not to report mean confidence rates for incorrect and correct decisions, because these rates are both poor predictors of the evaluated rates of TA correct, TA incorrect, TP correct and TP incorrect. Future research needs to place greater value on the TA and TP confidence rates separately for correct and incorrect decisions, before using the difference between mean correct and mean incorrect confidence rates to draw conclusions about an A-C relationship (as it was calculated in e.g. Read & Craik, 1995).

### **The value of mean confidence rates for TA and TP line-up conditions**

To evaluate whether mean confidence rates of the variable '*line-up*' are informative, again different confidence rates reported by Yarmey (1986) were re-evaluated (see table 2). The difference of TA line-up between correct and incorrect decisions is noteworthy. Participants who falsely identified a voice in the TA line-up, were far more certain about their decision ( $C = 3.75$ ), than participants who correctly rejected a voice ( $C = 1.65$ ). The report of mean confidence rates of TA line-up gives no information about these very different confidence levels, and therefore it is not of great value to use as a predictor of TA correct and TA incorrect confidence rates. The same conclusion holds for the mean confidence rate of the TP line-up, which is a poor predictor of TP correct and TP incorrect confidence rates. This results in a conclusion very similar to the one made in the previous section. Again, it can be concluded that future research needs to place greater value on the confidence rates of correct and incorrect confidence rates separately for TA and TP line-ups, before using the difference between mean TA and mean TP line-ups rates to draw conclusions about an A-C relationship.

### **The importance of splitting up incorrect decisions in false alarms and misses**

A remarkable aspect of Yarmey's study is the fact that confidence rates for misses (incorrect rejection in TP line-up) are reported separately from false alarms. In the other studies (Pickel, et al., 2003; Read & Craik, 1995; Stevenage, et al., 2011; Van Wallendaal, et al., 1994), confidence rates of these two groups were taken together, as incorrect decisions. However, as can be seen in table 2, the division of these two groups seems useful, because the confidence rates between the false alarm group and misses group (within TP line-up), are quite different (respectively,  $C = 3.5$  and  $1.4$ ). By splitting up these two groups, more information is obtained about the confidence levels that people have in different kinds of decisions (choosing or rejecting). The mean confidence rate of incorrect decisions, when false alarms and misses are taken together,  $C = 3.1$  (as it was calculated in Pickel, et al., 2003; Read & Craik, 1995; Van Wallendaal, et al., 1994), does not give this information, because it underestimates confidence rates of false alarms ( $C = 3.63$ ) and overestimates confidence rates of misses ( $C = 1.4$ ). Furthermore, this mean confidence rate of  $3.1$  underestimates the confidence rate for participants in the TA condition ( $C = 3.75$ ) and overestimates the confidence rate for participants in the TP condition ( $C = 2.45$ ).

The same effect is found to be true in the evaluation of the mean confidence



rate of the TP line-up condition (see table 2). Therefore, it can be concluded that splitting up incorrect decisions to false alarms and misses would also have been better in the other studies (Pickel, et al., 2003; Read & Craik, 1995; Stevenage, et al., 2011; Van Wallendael, et al., 1994) and would give a more detailed view of participants' confidence in their decisions. Another reason why reporting misses and false alarms separately is important, is because in reality it is less serious when a perpetrator is accidentally missed, than when an innocent person is falsely blamed. Moreover, since the confidence rates are much lower for misses than for false alarms, this emphasizes why it is worrisome to rely on these rates when using them to infer accuracy.

### **Conclusion correctness and line-up**

Altogether, it can be concluded that the mean confidence rates of both '*correctness*' and '*line-up*' are poor predictors of mean confidence rates. Because these mean rates could give a distorted view of the situation, it is important not to use them in the assessment of a relationship between confidence and accuracy. Although the study of Yarmey (1986) was the only study that reported information about the specific confidence rates between conditions, these rates were so distinctive that this conclusion could be made in this review.

Furthermore, it is very remarkable and predictable, that participants who chose a voice (hit or false alarm) were very similar in their certainty level. They all indicated their certainty level very high (around 3.6). On the other hand, participants who rejected a line-up (correct rejection or miss) indicated much lower certainty levels (around 1.5). In general, this indicates that participants are not very certain about a rejection, but are more certain about an identification. This fact seems to fit a statement that was made by Loftus (1979) (as cited in Van Wallendael, et al., 1994) who believes that participants have a tendency to make an identification instead of rejecting a line-up. Loftus (1979) explained this by the participant's belief that no test would be conducted, unless there was a reason for it. As a consequence, this mindset causes participants to not feel very certain about a rejection of a line-up, because it conflicts with their own thoughts. Because of this finding the previous conclusion, about not using mean confidence rates of correct, incorrect, TA and TP to assess the existence of a relationship between accuracy and confidence, is emphasized.

### **SUMMARY AND FUTURE DIRECTIONS**

This review discussed several issues that are related to the accuracy-confidence (A-C) relationship. It was investigated to what extent various measures of confidence could have an influence on the A-C relationship. This was evaluated by discussing different issues. First, two different measurements of confidence were evaluated, the discrimination method and the absolute accuracy method, which differed in the moment of measuring confidence of participants in their decisions. Several studies were discussed, to investigate if these two different measurements led to

different results in A-C relationships (Kerstholt, et al., 2004, 2006; McAllister, Dale, Bregman, et al., 1993; Orchard & Yarmey, 1995; Philippon, et al., 2007; Read & Craik, 1995; Saslove & Yarmey, 1980; Stevenage, et al., 2011; Van Wallendaël, et al., 1994; Yarmey, 1986). However, no obvious differences were found.

Studies that used the discrimination method found very inconsistent results; some found an A-C relationship, whereas others did not. On the other hand, studies using the absolute accuracy methods, found no or weak relationships. It is remarkable that all discussed studies differed in many aspects, for instance in the number of voices in the line-up. Since this could influence the results, this made the comparison very difficult (Bull & Clifford, 1984). Another point that made a comparison very difficult was the difference in confidence scales used. Some studies used a 3-point scale (Saslove & Yarmey, 1980), whereas others used a 4-, 5- or even a 7- point scale (e.g. Orchard & Yarmey, 1995; Stevenage, et al., 2011; Yarmey, 1986). These different scales could also influence the results, since a scale from five to ten was found to be better than a shorter scale (Forzano & Gravetter, 2009).

Finally, the fact that most experiments do not clearly specify how they calculate the confidence rates that are used in the A-C calculation, especially when they use the discrimination method, made a comparison very difficult. Due to these differences between the discussed studies, I conclude that future research is needed to investigate if these two different confidence measurements lead to different results about the A-C relationship. However, until this difference is clearly investigated, it is advised to use the discrimination method, because this method gives more information about the memory processes used by people to make a decision.

In the second section, the potential risk of confounding in confidence rates (C) was discussed. Most studies measuring the accuracy-confidence relationship use mean confidence rates of all participants to assess this relation (mean C within the 'correctness' of a decision and mean C within the different 'line-up' conditions). However, after re-evaluating several mean confidence rates reported by the study of Yarmey (1986), it was concluded that these mean confidence rates are no reliable predictors of specific confidence rates for participants who accurately reject a line-up (correct rejection), participants who accurately identify a voice (hit), participants who inaccurately reject a line-up (miss), and participants who inaccurately identify a voice (false alarm). Mean confidence rates of correct decisions (hits, and correct rejections) were found to be poor predictors of these specific confidence rates. The same was found to be true for mean confidence rates of incorrect decisions, target present, and target absent line-ups. Only two confidence rates were found to be predictable, namely the mean confidence rate for people who (correctly or incorrectly) chose a voice, which was very high ( $C \approx 3.6$ ), and the mean confidence rate for people who (correctly or incorrectly) rejected a line-up, which was very low ( $C \approx 1.5$ ) (Yarmey, 1986). Therefore, it was concluded that mean rates should not be used in calculation of the A-C relationship, but that it is better to evaluate specific confidence rates to get information about people's decisions. In reality it remains, however, very difficult to implement all findings about A-C relationships, because in reality the real truth is never known. Therefore, it remains tricky to use confidence rates in the assessment of accuracy.

The conclusions of this review should be interpreted with caution, because of the low number of papers that could be included in the discussion. Earwitness testimony has been investigated less than eyewitness testimony, and so overall fewer papers could be found regarding this issue. Moreover, because issues were discussed and re-evaluated for which specific information was needed, only a few studies were useful to include in this review. However, despite the fact that this limitation causes the conclusions to be more difficult to generalise, it emphasizes that much more research is needed in this area. Not until more systematic research on the discussed issues is conducted, it is crucial to be careful in reporting whether a relationship between accuracy and confidence in earwitnesses exists or not in research on earwitness testimony. In this way, this review may serve as an eye opener for future research about specific issues related to earwitness testimony.

## REFERENCES

- Boydell, C. A., & Read, J. D. (2011). Accuracy of and confidence in mock jailhouse informants' recall of criminal accounts. *Applied Cognitive Psychology, 25*, 255-264.
- Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 92 - 123). New York: Cambridge University Press.
- Forzano, L. A. B., & Gravetter, F. J. (Eds.). (2009). *Research Methods for the Behavioral Sciences*. Belmont, CA: Wadsworth.
- Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, A. P. A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology, 18*, 327-336.
- Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology, 20*, 187-197.
- McAllister, H. A., Dale, R. H. I., Bregman, N. J., McCabe, A., & Cotton, C. R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology, 14*, 161-170.
- McAllister, H. A., Dale, R. H. I., & Keay, C. E. (1993). Effects of lineup modality on witness credibility. *The Journal of Social Psychology, 133*, 365-376.
- Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011). Typicality effects on memory for voice: Implications for earwitness testimony. *Applied Cognitive Psychology, 25*, 29-34.
- Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9*, 249-260.
- Philippou, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology, 21*, 539-550.
- Pickel, K., French, T., & Betts, J. (2003). A cross-modal weapon focus effect: The influence of a weapon's presence on memory for auditory information. *Memory, 11*, 277-292.
- Quinlivan, D. S., Neuschatz, J. S., Jimenez, A., Cling, A. D., Douglass, A. B., & Goodsell, C. A. (2009). Do prophylactics prevent inflation? Post-identification feedback and the effectiveness of procedures to protect against confidence-inflation in earwitnesses. *Law and Human Behavior, 33*, 111-121.

- Read, D., & Craik, F. I. M. (1995). Earwitness identification: Some influences on voice recognition. *Journal of Experimental Psychology: Applied*, 1, 6-18.
- Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology*, 65, 111-116.
- Stevenage, S. V., Howland, A., & Tippelt, A. (2011). Interference in eyewitness and earwitness recognition. *Applied Cognitive Psychology*, 25, 112-118.
- Van Wallendael, L. R., Surace, A., Parsons, D. H., & Brown, M. (1994). 'Earwitness' voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology*, 8, 661-677.
- Yarmey, A. D. (1986). Verbal, visual, and voice identification of a rape suspect under different levels of illumination. *Journal of Applied Psychology*, 71, 363-370.
- Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, 1, 792-816.

# What's the problem with free will?

## REVIEW

The debate about the existence of free will is often referred to as “the free will problem”. However, this essay intends to illustrate that “the free will problem” may not exist after all but that discussions about free will suffer from a multitude of problems. In particular, definitions and theories lack consensus about the nature of free will and are often too general. Support for arguments is often provided by citing neuroscientific studies despite the fact that these suffer from methodological limitations and allow multiple interpretations depending on the very definition of free will. Furthermore there is the risk of undermining the belief in moral responsibility in society by premature public discussions.

“The” problem of free will is therefore, according to this essay, that inter- and intra-disciplinary teamwork in this debate is not optimal yet. It is concluded that a consensus on a definition about free will is a prerequisite for advances in this field of science.

**Keywords:** free will, free won't, moral responsibility, neuroscience, philosophy

Marianne Drabek; research master student Cognitive Neuroscience  
Maastricht University, Maastricht, the Netherlands

---

[m.drabek@student.maastrichtuniversity.nl](mailto:m.drabek@student.maastrichtuniversity.nl)

## INTRODUCTION

The existence of free will has been debated amongst philosophers for decades. This debate has drastic implications for every individual as it is most natural for humans to believe and rely on the assumption that everyone is the master of one's own actions. Many (if not most) individuals would feel deeply uncomfortable if this is not the case since it would mean that they perform actions and express wishes that they do not generate consciously, not unlike a puppet.

At the same time the nonexistence of free will is often claimed to be a threat to society and to require fundamental changes in its penal system since wrongdoers cannot be held accountable for their actions if they are not masters of their will. 'Unsoundness of mind' is an important concept in the legal system but denying the existence of free will altogether implies that nobody could be held responsible or punishable for anything. The mere thought that crimes like murder or rape would not even be attempted to be prosecuted is dreadful.

The tremendous potential for conflict that this topic holds for society is mirrored by the eagerness to discuss this matter: the search term "free will" yields 19,000,000 hits in google, 3,590,000 hits in google books, 263,000 hits in google scholar, and 4,741 hits in ScienceDirect (19.04.2013). While the chosen databases are only examples and these numbers are only approximate, the order of hits is informative enough to imagine the dimensions of this debate.

In the last 30 years this debate tightened as neuroscientific experiments are often cited from critics as proof for the claim of free will being an illusion. Yet, what exactly is being cited and what did these neuroscientific experiments show?

This paper aims to elucidate some problems associated with the *debate* about free will and tries to assess their potential. Is the non-existence of free will supported empirically? What does the debate mean to society? And more importantly: What is being debated, in other words: what exactly is free will? Some answers will be given in this essay, in particular by showing that the answer to the last question is far from easy.

The first section reviews some definitions of free will and sketches some considerations about defining free will. This is followed by reviewing neuroscientific experiments with regard to free will and their limitations. The third section discusses possible issues for society arising from the current debate. Lastly, a summary with conclusions completes the essay.

## WHAT EXACTLY IS FREE WILL? PROBLEMS WITH DEFINITIONS AND THEORIES

While the meaning of free will seems intuitive for most humans, it is not easy to give an exact definition, as the following section will show. Philosophers often discuss free will with regard to moral responsibility and its existence, mostly by referring to common views including (in)compatibilism, (in)determinism, libertarianism and subtypes thereof (for an overview see Pockett, 2007). Yet, few philosophers describe the *underlying nature* of free will. However, this is important, because definitions determine how arguments are used and what exactly it is that is being discussed; many times this seems not to be taken into account. Neither intending to summarize all possible opinions nor to argue for or against any perspective, a few accounts on free will are described in this section to demonstrate how different assumptions about the nature of free will influence the overall debate.

### Problems due to different assumptions about free will

Some philosophers, the most famous being Descartes (see The sixth Meditation, translated by Moriarty, 2008), separate free will from matter and describe it as a metaphysical force from which thoughts and actions originate. This dualistic view is clearly fairly mysterious, neither possible to prove, nor easy to disprove completely. Others, like Wegner, describe free will as a feeling, sensation or emotion, “not unlike happiness or sadness...” (Wegner, 2004, p.658). This way, Wegner argues, acts are attributed to the self in a post-hoc manner which makes free will illusory. The logic is that if the *thought* is *consistent* with and occurs before the movement and if there is no other “obvious cause for the movement [t]hese features imply causality, that the thought led to the movement“ (Hallett, 2007, p.1182).

While he does not distinguish properly between feeling, sensation, emotion, or percept, Gray Hardcastle illustrates the importance of exact wording. She commented that “[o]ur sensation ....may get it wrong once in a while; it may get it wrong lots of times. Nevertheless, the sensation is reflecting something real, as real as our bodies’ need for nutrients. The important question is what exactly is that sensation reflecting.” (Gray Hardcastle, 2004, p.663). When looking at free will this way, this definition goes beyond mere perception; mistakes may happen but there is a ‘real’ basis to the sensation. Already such a relatively small difference (compared to viewing free will only as a perception) reshapes arguments about free will.

Yet others, mostly compatibilists like Frankfurt, link free will to decisions or choices and thus ability (for overview of compatibilism see Campbell, 2011). It is important to distinguish this account from the other in that decisions necessarily precede actions, a feature that has important implications for arguments about free will’s existence. Already these few examples hint at the degree of variation that a definition on free will can include: In comparison to the meta-physical force account, viewing free will as a percept or emotion offers more loopholes, e.g., that a feeling may be mistaken in certain situations and also explain some of the function that free will may have.

### Problems due to missing limitations in definitions

In fact, the inclusion of loopholes and limits make definitions resistant against arguments about exceptional cases and thus make them generally valid (i.e. the definition cannot be proven to be wrong by such arguments because exceptions or limits are specified). Unfortunately, the range of cases to which a definition can be applied is rarely discussed although this is a crucial component: Is it necessarily a (an entirely) conscious process and what process is defined to represent free will: “The earliest initiation of action process” or “the choice or selection of a *specific* action” (Haggard & Libet, 2001, abstract. Bold print in original instead of italics)?

When does free will develop (or does everyone possess it from birth)? Does everyone possess free will (e.g., children, easily manipulable individuals, cognitively impaired or mentally disordered)? Under what circumstances is one capable to exert free will concerning specific decisions (e.g., being in love, exceptional emotional situations, physical conditions like hunger or sleep lack, substance abuse, peer pressure, brain tumors)?



Especially with regard to definitions involving decisions, it may be required to impose certain restrictions otherwise discussion may be strangled by arguments about physical needs (e.g., mild hunger when shopping interfering with choices) and impulsivity. Restricting free will to higher-order processes would help to clarify discussion grounds and avoid debates about ambiguous situations. At the same time, even such a restriction may be too vague because emotions and context influences can never be completely left out of any decision. Thus, it is needed to specify the type of decisions in question and the degree of being informed about the different choices and corresponding consequences.

When free will is described, it is often briefly defined to involve no coercion or constraints. However, decisions may not necessarily reflect the will of an individual because of the limited choice possibilities of a situation. Examples demonstrating that a person's will may not match any of the possibilities in a particular situation are plentiful (even if budget is unlimited, certain desired things may not be available). This makes it questionable that true "no-constraints situations" exist in a nonutopian world and thus precision is needed in such statements. Would it be better to speak of a person's 'wishes' or 'desires' rather than 'decisions'? This could be an entire discussion on its own; the example is only mentioned here to illustrate again the influence of phrasing. Important is, though, that 'free' should not be confounded with freedom.

The previous point was raised about external coercion (or constraining) factors. What about internal factors: Could one not argue that past experience or physical needs already dictate? If so, one could object that a person will never be free because of influences of the past and habits or attitudes that resulted from it (e.g., eating sweet things since childhood); but then exactly how free is 'free'? One could argue that attitudes are part of the person's personality rather than internal coercion factors since the person may have actively decided to adopt a particular stance (e.g., 'I do not buy this product because I'm vegetarian out of principle...'); yet again these conscious decisions and attitudes may have developed due to past experience. On the other hand, some consistency in a person's decisions is expected (since they are generated by the same person) but at what point is the line crossed to favor determinism?

While certainly leading, a person may still decide differently from what attitude or habits dictate (e.g., '...but this time I'll make an exception'). Is the ability to veto enough to prove that there is a 'real choice'? Some philosophers and scientists think so since they defined free will by quite the opposite, a 'free won't' (e.g., Obhi & Haggard, 2004) instead of an active decision (free will).

This shows that many different cases are needed for thought experiments to develop a waterproof definition instead of a general statement. Certainly many of these questions are bound to be discussed but these jigsaw elements are spread in the literature, not agreed on and often omitted all together.

As a conclusion, free will requires a complex definition because the very nature of free will and possible limitations need to be addressed. This may be tedious but possible if constructive discussions are held. Sufficient time should be taken since

fast and premature definitions lead to misunderstandings, circular arguments and will only hamper progress.

Moreover, the very definition of free will is a basis for experimental rationales and interpretations, as the next section will show, which makes it clear that a consensus on the definition is a prerequisite to advance this field of science with empirical experiments.

## PROBLEMS OF NEUROSCIENCE WITH EMPIRICAL INQUIRIES ABOUT FREE WILL

Empirical results are always valid only in the context of a specific definition and with respect to their limitations. While this seems so obvious that it is hardly worth mentioning, limitations of interpretations are often left aside in debates about free will by citing only what are believed to be the main findings. This section explains some of the methodological weaknesses of neuroscientific experiments in the context of free will. In particular, the ‘classic Libet experiment’ will be discussed in this section followed by some theoretical and methodological considerations against the common interpretation of this and similar results.

### **The ‘classic Libet experiment’**

In almost any introductory part in articles about free will, the so called “Libet experiment” (Libet et al., 1983) is summarized as the first and cornerstone of neuroscientific evidence against free will. Using electroencephalographic (EEG) recordings from subjects, who were told to make a voluntary key press at a time of their choice, Libet and colleagues showed readiness potentials 1 second prior to the execution of the movement. These potentials occurred prior to the subject’s own awareness of his or her decision. To show this temporal component of their awareness, Libet’s subjects were asked to observe a rotating clock and to note the clock’s position when they were aware of their urge to press either of the two keys with either left or right hand. They found that the readiness potential was indicative of the choice and occurred prior to the time of becoming aware of the decision, as indicated by the reported clock’s position. This has often been taken as evidence that the subject’s choice was determined by something else than one’s own free will.

### **Problems with the interpretation of Libet’s findings**

Similar experiments have been carried out using single-neuron recordings and fMRI but this experiment is by far the most cited paper in these discussions. Yet, comparatively few articles mention the limitations or doubts against the common interpretation of his data. Interestingly, Libet himself did not claim to have proven the non- existence of free will. On the contrary, he explains that the movements of the subjects could be initiated subconsciously and that while such a decision may be generated without the subject being aware, the mind still possesses a veto power to not execute the act (Libet, 1999). In this respect Libet seems to subscribe to a

“free won’t”, emphasizing that a definition on free will should not be made light-hearted. However, even if we hold on to the definition of free will rather than the ‘will to do otherwise’ as it is also referred to, there are more objections against hasty conclusions.

There is the possibility that the activity measured (either by EEG or fMRI or electrophysiologically) *reflects preparatory activities for the decision making or even the decision process itself* (Trevena & Miller, 2002). This process may again consist of both subconscious and conscious components, all of which makes temporal data (e.g., by EEG) uninterpretable in terms of what activity represents free will or what is predictive of free will.

In fact, finding biological components or mechanisms of a decision via EEG or fMRI or single-cell recordings does not even necessarily mean that dualistic beliefs need to be abandoned because it is known (e.g., from neurofeedback) that the *mind is capable of influencing brain activity* (Stier, 2011, p. 990. translation). Does the mind influence brain activity or is it the other way round? Due to the possible entanglement of the various (sub)conscious components in a complex decision process, this seems to resemble a hen-egg question. Moreover while results may be found that explain certain features, Roskies argues that neuroscience is not in the position to *prove that free will is reduced to [one] mechanism only* (Roskies, 2006, p.421).

This shows that there are already sufficient extensive theoretical considerations which should at least caution against absolute claims (that free will is an illusion) without the need to refer to methodological weaknesses; a selection of these will nevertheless be mentioned in the following.

One potential methodological weakness is the choice of tasks in these experiments. As Haggard notes, participants are usually asked to make decisions about personally irrelevant or meaningless choices, such as right or left key presses. Thus, there is “generally no reason or value that motivates the participant to choose one action over another” (Haggard, 2008, p. 934), which may mean that there was no need for the participant to have any free will in these moments. It follows that tasks may need to involve higher-cognitive decisions; a point that is also relevant when reviewing animal studies on free will. Nevertheless, even when the tasks get more complex or closer to everyday life, the argument may remain that the situation is a laboratory-one and therefore too well controlled by the subject (Stier, 2011), who, in addition, may make (unconscious) efforts to please experimenters.

Further, it is questionable whether the temporal components measured are actually reflecting meaningful temporal order. Trevena and Miller who conducted an experiment similar to Libet’s, acknowledge that averaging over trials may cause *smearing effects in readiness potentials* which may completely misrepresent temporal order (Trevena, & Miller, 2002). Further inaccuracy may be introduced by *perception difficulties*, i.e., delay which is introduced when the person indicates the time point of the decision (Trevena & Miller, 2002). Similarly, there has been research, showing that the “subjective present is actually slightly in the real past” (Hallett, 2007, p.4).

On top of that, it has to be noted that the temporal resolution of the methods

to study brain activity may not be close enough to the resolution needed. This is sure for fMRI which has a temporal resolution in the order of several seconds. However, despite the fact that the temporal resolution of EEG in milliseconds is fairly good one cannot be certain that it suffices for absolute claims (Stier, 2011).

As it stands, caution is recommended when using neuroscientific data in arguments about free will. For fairness it should be mentioned that there is also empirical work aiming to demonstrate the existence of free will; it suffers from the same limitations, though: definitions, theoretical considerations and methodological weaknesses (review by Stier, 2011). The best methods currently available may not provide satisfactory quality for the problem studied. Perhaps it is even impossible to create an experimental design that is complex enough and immune to methodological inaccuracies. Is neuroscience able to advance the free will discussions at all? Even though neuroscience did not achieve reliable evidence for or against free will yet, this research has already moved a mountain by initiating interdisciplinary talk. Hopefully this will lead to refined definitions and theories at the least. The next section will briefly show what the debate means to society until the debate has come to a consensus.

## PROBLEMS EVOLVING FOR SOCIETY FROM FREE-WILL DISCUSSIONS

Without a doubt, the discussion about free will may have drastic implications not only for the common justice system but also for every individual. While this problem has the potential to be tremendous, as outlined in the introduction, several researchers and philosophers claim that it may not be a problem after all- at least not for anyone who is not a dualist (Hallett, 2007). Hallett for example argues that “[a] person’s brain is clearly fully responsible, and always responsible, for the person’s behavior” (Hallett, 2007, p.1189). Adherents of a ‘free won’t’ could also argue a person could have used their veto power to prevent the action from execution (Libet, 1999), which would still make them responsible. Again, this discussion depends to some degree on a definition, but regardless of the exact arguments there seems to be some agreement that the legal system should not be changed. If so, is the conflict solved then?

The answer is no, since there may still exist a problem for the society arising from the discussions about free will and determinism. When Vohns and Schooler primed participants with texts on determinism, their participants cheated considerably more often and thus were led to behave immorally (Vohns & Schooler, 2008). Baumeister and colleagues found that subjects behave more aggressive and less social after being primed that free will may not exist (Baumeister et al., 2009).

These and similar data should be a warning to the scientific community that there is the potential for great harm for society when careless phrases and incomplete discussions (until a consensus in science is reached) reach a non-critical audience. Some individuals in such an audience may not assess such discussions correctly and may make use of any excuse to abandon responsibility for misdeeds.

## 'THE' PROBLEM OF FREE WILL? SUMMARY AND CONCLUSION

This last section summarizes and concludes what "the" problem of free will is (according to the author of this essay). The first section focused on the problems of various definitions and underlying assumptions and illustrated what implications these discrepancies have. Sufficient time should be taken to work on these issues since fast and premature definitions lead to misunderstandings and will only hamper progress. Moreover, the very definition of free will is a basis for experimental rationales, which makes it clear that a consensus on the definition is of outmost importance in order to advance empirically.

The next section described some of the concerns that have been voiced against the common interpretation of Libet's experiments and similar work. There have been considerable theoretical and methodological concerns such that current empirical data are neither suitable to fully support contra nor pro free will arguments at this stage. Such arguments include for example the entanglement of conscious and subconscious components in the decision process or the temporal interpretation of data, or the type of tasks in these experiments. The problem is thus that neuroscience has yet to come up with adequate strategies and methods as to how free will can be studied.

The problem illustrated in the last section concerns problems for society from debates about free will. The illusion of free will (should it be proven) would perhaps not cause problems for legal responsibility. Instead there is an immediate and observable threat in incomplete discussions and premature conclusions, because they may tempt a less critical audience to abandon the concept of moral responsibility, as has been shown to be the case in several studies.

Given that the problem of free will is tightly linked to its definitions, these problems could be expected to be handled after sufficient discussions. However, a prerequisite for this is patience, well-working interdisciplinary teamwork and mutual respect of philosophers and neuroscientists as equal partners in this discussion. In addition, good scientific practice includes that results are always reported and cited with regard to their respective limitations. If this succeeds, eventually, a consensus on a definition will be reached and empirical advances can be built up on this. Until then, necessary caution should accompany discussions to protect society.

### **Acknowledgments**

I want to thank my dad for being a passionate listener and the reviewers for constructive comments.

## REFERENCES

- Baumeister, R.F., Masicampo, E.J., DeWall, C.N. (2009). Prosocial Benefits of Feeling Free: Disbelief in Free Will Increases Aggression and Reduces Helpfulness. *Personality and Social Psychology Bulletin*, 35(2), 260-268. doi: 10.1177/0146167208327217.
- Campbell, J.K (2011). Free Will. Cambridge, *Polity Press*. pp. 86.
- Descartes, R. (2008). Meditations on First Philosophy with Selections from the Objections and Replies. A new translation by Michael Moriarty. Oxford World's Classics. New York, *Oxford University Press*.
- Gray Hardcastle, V. (2004). The elusive illusion of sensation. [Comment on Wegner]. *Behavioral and Brain Sciences*.
- Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, 9(12), 934-946.
- Haggard, P., & Libet, B. (2001). Conscious intention and brain activity. *Journal of Consciousness Studies*, 8(11), 47-63.
- Hallett, M. (2007). Volitional control of movement: The physiology of free will. *Clinical Neurophysiology*, 118(6), 1179-1192.
- Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies*, 6(8-9), pp. 47-57.
- Libet, B., Gleason, A.C, Wright, E.W., & Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106, 623-642.
- Obhi, S., & Haggard, P. (2004). Free will and free won't. Motor activity in the brain precedes our awareness of the intention to move, so how is it that we perceive control? *American Scientist*, 92(4), 358-365.
- Pockett, S. (2007). The Concept of Free Will: Philosophy, Neuroscience and the Law. *Behavioral sciences & the law*, 25(2), 281-293.
- Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences*, 10(9), 419-423.
- Stier, M. (2011). Hirnforschung pro Willensfreiheit. Was ist dran? *Nervenheilkunde*, 30, 987-991.
- Trevena, J.A., Miller, J. (2002). Cortical Movement Preparation before and after a Conscious Decision to Move. *Consciousness and Cognition*, 11, 162-190. doi:10.1006/ccog.2002.0548 .
- Vohns, K.D., & Schooler, J.W. (2008). The Value of Believing in Free Will. *Psychological Science*, 19 (1), 49-54. doi: 10.1111/j.1467-9280.2008.02045.x
- Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral and Brain Sciences*, 27(05), 649-659.

# ‘Walking through the hospital’, priming tolerance to enhance the approaching behavior between healthy individuals and patients suffering from psychoses

## ORIGINAL PAPER

Research has shown that priming works, but most evidence comes from laboratory situations and may not be applicable to situations in daily life. The present study used priming for the enhancement of the quality of approaching behavior between patients suffering from psychoses and healthy young individuals. Twenty-one healthy undergraduate students (age: 18-25) were recruited. Interpersonal distance was used as an indicator for approaching behavior. Using a virtual reality setting, it was tested whether the participants keep less distance towards psychotic patients after being primed with a sentence scrambling task involving words associated with tolerance. It was hypothesized that priming of tolerance can elicit more empathy towards mentally ill. The data reveal that this method of priming has no measurable influence on approaching behavior in situations in which people encounter mentally diseased suffering from psychoses. Nevertheless, this needs further investigation. Priming for social purposes could be a convenient and supportive way to improve people’s attitude towards mental illnesses. **Keywords:** social distance; stigma; psychosis; tolerance priming; virtual reality.

Lisa Hubrach; bachelor student year 3  
Kiki Nap; bachelor student year 3  
Maastricht University, Maastricht, the Netherlands

---

[l.hubrach@student.maastrichtuniversity.nl](mailto:l.hubrach@student.maastrichtuniversity.nl)  
[k.nap@student.maastrichtuniversity.nl](mailto:k.nap@student.maastrichtuniversity.nl)



## INTRODUCTION

The movie “Black Swan” is a popular movie dealing with mental illnesses. In the movie, Natalie Portman (alias Nina Sayers) experiences the consequences of a psychosis. As the watcher of the movie, you will see how two worlds can exist in one person. This movie made a contribution to the development of making mental diseases more familiar to a broad range of people by showing them the daily life of someone having a psychosis.

According to a general investigation of Vaughan and Hansen (2004) about mental illnesses, the society got used to the occurrence of mental diseases over the last decade. A lot of people learned that those disorders should be taken as seriously as physical ones. The discussion about mental illnesses and its classification is more and more expanding from psychologists to the general population, resulting in more attention to the tabooing of mental illness and the end of this taboo (Vaughan & Hansen, 2004).

In spite of those positive findings, people tend to stigmatize mentally ill persons. Sickened persons are treated differently and lose support by their caregivers (Stout, Villegas & Jennings, 2004). It has been shown, for example, that people tend to keep more distance, both emotionally and spatially, towards persons with abnormal behavior than to persons with normal behavior (Lauber, Nordt, Falcato & Rössler, 2004). The distance a person keeps towards the partner he or she is interacting with, has been shown to be a good indicator whether a person feels comfortable during the interaction (Bailenson, Blasovic, Beall & Loomis, 2001).

A lot of different attempts have been made to reduce the skepticism against mentally disordered (e.g. Vaughan & Hansen, 2004). For instance, research is conducted to investigate the shortcomings of media usage in stereotype reduction. Mainly, media are challenged on how to permeate to the consciousness and attitudes of its recipient (Stout, Villegas & Jennings, 2004).

A more subtle and experimentally measurable way than media campaigns to reduce discrimination and to change behavior against mentally ill persons may be priming. Priming is a technique to implicitly influence the thinking and behavior of persons. Through unconscious cues, certain aspects or schemes within the mind are made more salient (Bargh, Chen & Burrows, 1996). When those aspects are more accessible, they can guide the person's behavior in a certain direction. In that way it may be possible to influence a person to act in a specific manner. The active self account by Wheeler provides a theoretical basis explaining the mechanism of priming (Wheeler, DeMarree & Petty, 2007). According to this theoretical framework, the prime activates a certain facet of the self, which then leads to a particular kind of behavior. In a study of Hansen & Wänke (2009) the facet “professor” through priming of the professor stereotype was activated. This increased the accessibility of aspects from the self related to that facet. When the “professor-facet” was triggered, the parts of the self connected to intelligence and performance were activated. This led to better performance during a knowledge test (Hansen & Wänke 2009).

Priming was and is frequently used to influence behavior in several ways. Other experimental studies used it to influence motor behavior of people (i.e. walking more slowly after priming the elderly stereotype; e.g. Bargh, Chen & Burrows, 1996)

or the judgment of neutral faces through odors (Li, Moallem, Paller & Gottfried, 2007), and even donation behavior (Lamy, Fischer-Lokou & Guéguen, 2012).

The present study aimed to apply the framework of Wheeler et al. (2007) and the findings from previous studies in the clinical sector. In this study, we wanted to investigate whether priming is a valuable way to reduce stigmatization of mentally sickened persons. Using priming, it might be possible to provide social guidance on how to react in situations with mentally diseased persons. A psychosis was chosen as mental illness. Although all mental diseases tend to utter themselves very differently, we thought that a psychosis could be the most striking example of a mental illness causing the relatives to be insecure about how to interact with such persons. A psychosis is characterized by negative and positive symptoms. Negative symptoms involve the reduction of emotionality, stereotyped thinking and less perspective taking (Piskulic & Addington, 2011). In contrast to that, positive symptoms are experiences which are added to the normal human functioning. Positive symptoms include hallucinations and delusions. For instance, the patient always feels persecuted (DSM IV-TR, 2000).

Based on the active self account of Wheeler, DeMarree & Petty (2007), we attempted to activate the tolerant facet of the self to influence the approaching behavior to mentally disordered persons. Because of the fact that there is a link between the distance a person keeps and the comfortableness during an interaction (Lauber, Nordt, Falcato & Rössler, 2004), we inferred that less distance is a sign of reduced stigmatization.

A virtual reality (VR) experience was used to investigate whether the approaching behavior to a mentally diseased person was improved after a tolerance prime. A VR setting was used, because it has some advantages in comparison with other methods. First, it is possible to measure interpersonal distance in a scientific and precise manner. The participant was not aware of the fact that her/his distance to the virtual person was measured. In addition to that, the tracking system of the virtual reality laboratory is able to measure very precisely using the 16 cameras it is equipped with. A second advantage is that the situation is standardized across participants because all participants encountered the same persons with exactly the same behavior. If we asked a simulation patient to be our patient, it could have happened that he/she would have interacted differently across participants.

Wheeler & Berger (2007) found that equal primes can have different effects depending on the personality structure. Because of that, the personal tolerance level of the participant was taken into consideration. The question was whether the personal tolerance level increases or decreases the effect of the tolerance prime.

It was hypothesized that priming of tolerance will activate a facet of the self concept, which improves the approaching behavior to mentally ill persons. This could be expressed by a decrement of distance. Moreover, it was hypothesized that the personal tolerance level would lead to a different effect of the prime. If the tolerance level already is high, then the facet related to tolerance already is active. Because of that, it could be that the prime does not add any new information to the active facets. Subsequently, it could be that the prime has no effect or that highly tolerant persons show a different priming effect than low tolerant participants.

## METHODS

### Participants

After approval of the study by the ethical committee of Maastricht University's Faculty of Psychology and Neuroscience, a total of 21 participants aged between 18 and 25 years (9 men, 12 women) were recruited. Seventeen participants were psychology students of Maastricht University and 4 were students of other faculties of the same university. Participants were admitted to this study if they did not experience seasickness to prevent motion sickness provoked by the virtual reality experience (Howarth & Costello, 1997). Two participants were excluded from the experiment; one due to lost VR data and one because of poor perception (i.e. the participant did not notice it when the mental or physical condition of the patient switched). Psychology students amongst the admitted participants were rewarded with 1,5 participation point while the remaining participants participated on a voluntary basis.

### Materials

#### *Tolerance level questionnaire*

We used a paper version of the questionnaire of Devine and Plant (1998) to measure tolerance level. Considerable changes were made in the questions, because the questionnaire of Devine and Plant only tested prejudice against black and white people and we wanted to focus also on other prejudiced groups (e.g. homosexuals, mentally ill people and skinheads). It therefore was decided to replace for example 'black people' by 'mentally ill people'. Participants were tricked into thinking that social ability was measured instead of tolerance level. Items were added and the questionnaire was labeled a 'questionnaire to measure social ability'. The reason for this distraction was to decrease the opportunity the participants got to know the actual goal of the study, which could influence the results in a negative way. The items added included questions about culture and relationships (see table 1). These distractors were not included in our analysis. Participants used a 9 point scale to rate each statement (ranging from 1: strongly disagree to 9: strongly agree).

**Table 1:** The modified tolerance level questionnaire (Plant & Devine, 1998).

- 
- |    |   |
|----|---|
| 1  | Because of today's politically correct standards, I try to appear non prejudiced toward black people.           |
| 2  | I try to hide any negative thoughts about mentally ill people in order to avoid negative reactions from others. |
| 3  | I attempt to act in nonprejudiced ways toward impaired people, because it is personally important to me.        |
| 4  | Because of my personal values, I believe that using stereotypes about homosexuals is wrong.                     |
| 5  | I am personally motivated by my beliefs to be nonprejudiced toward people who are different.                    |
| 6  | According to my personal values, using stereotypes about mentally ill persons is OK.                            |
| 7  | I attempt to appear nonprejudiced toward mentally diseased persons in order to avoid disapproval from others.   |
| 8  | Being nonprejudiced toward impaired people is important to my self concept.                                     |
| 9  | I try to act nonprejudiced toward impaired people because of pressure of others.                                |
| 10 | If I acted prejudiced toward homosexuals, I would be concerned that others would be angry with me.              |
- 

### *Sentence scrambling task*

Colleague students (n=17) were asked to write down twenty words they associated with tolerance. On this basis we sampled the most frequently chosen words for the priming condition. In total fifteen words out of 56 words were selected for the priming condition. One of the five words of each sentence scrambling task was 'highly tolerant', whereas the other words were neutral. One example of the words used in the priming condition is: 'respect, people, to think, public, women'. In this case we indicated respect as a highly tolerant word. The nonpriming condition consisted of neutral words like: 'blue, summer, sky, vacation, Italy'. The participant was asked to make a sentence out of these words. For example: This summer, I was on vacation in Italy and the sky was blue. The participant was asked to do this 15 times with different words. The participant received a paper (with the indicated words on it) and a pen and he or she had to make sentences out of the indicated words. The experimenter mentioned that it was important to make the sentences grammatically correct, to make the coverstory work well.

**Table 2:** The items of the word scrambling tasks (the bold words are indicated as ‘highly tolerant’).

	Priming condition	Non priming condition
1	respect, people, to think, public, women	blue , sky, summer, vacation, Italy
2	differences, to accept, men, dinner, cook	university, exam, book, library, quiet
3	election, to vote, liberal, democracy, room	car, color, price, kilometer, speed
4	students, unprejudiced, most, black, people	hair, black, red, blonde, short
5	open minded, journey, New Zealand, money, experience	party, drinking, friends, boyfriend, bed
6	mental ill, people, empathy, psychosis, help	airplane, train, car, bike, speed
7	children, roles, to play, tolerant, behavior	steak, fries, sugar, cola, broccoli
8	friends, to be, equal, sports, clothes	music, rhythm, voice, microphone, stage
9	to wait, train, clock/watch, patience, rain	Weather, sun, storm, rain, umbrella
10	art, colors, exposition, artist, interest	trousers, t-shirt, socks, shoes, blue
11	welcoming, host family, country, gift, special	book, story, film, scene, actor
12	Life, possibilities, opportunities, to choose, pleasure	life, optimism, pessimism, person, character
13	Facebook, media, contact, social, like	baby, diaper, milk, sweet, little
14	Christmas, together, tree, wine, fire	Christmas, tree, ginger bread, wine, fire
15	To follow, moral, important, to have, values	experiment, researcher, test, variable, statistics

### *The virtual environment*

The experiment was conducted at the immersive virtual environment research laboratory at the Faculty of Psychology and Neuroscience of Maastricht University, in a 5.7 by 7.8 meters sized room. The Immersive Virtual Environment Technique (IVET) setup contained sixteen cameras and an equal number of speakers. The cameras placed across the room accurately tracked the so called light emitting diodes (LEDs) located on the stereoscopic head-mounted display (HMD). Thus, using the

tracking of the LEDs, the participant's position and orientation were measured during the experiment. The HMD contains build-in screens that are placed in front of the eyes, displaying the graphics of the IVET rendered by a computer for each eye separately. A laptop was carried on the back gathering the participant's position and head-movement information which was rendered to the HMD, thereby creating the experience of being immersed within a 3D virtual environment.

### *Modeling of the immersive virtual environment*

The same type of immersive virtual hospital room environment was used as in the PhD project of Toppenberg (Toppenberg, Bos, Wigboldus, & Pryor, 2009-2014; papers in preparation). The facial editor software from FaceGen (2009) was used for creating the 30-year-old virtual male faces. Two types of patients were distinguished (both were lying in a bed): a patient with a mental disease (psychosis) and a patient with a somatic injury (broken leg). The patient with the psychosis was fixated to the bed with two blue bands. This patient looked very pale. The patient with the broken leg was lying in bed with his leg upwards (with his knee in a knee cast, so that the plaster was visible). Furthermore, this patient looked healthy. Both patients had nametags on their wrists and both were following the participants with their eyes. The hospital room was relatively empty; in the middle of the room a bed was standing and some apparatus were present. The room was light and spatial, because one window (with bus station view) was present.

### *Interpersonal distance*

The wrist was chosen as reference point to measure interpersonal distance (X, Z coordinates in VR; 0,75; 0,67). This reference point was consistent across all trials. Furthermore, it was chosen to measure from the foot of the bed ( $Z=0,5$ ), this was to make sure we only measured the distances that were relevant in our experiment.

## **Procedure**

After giving informed consent, participants were asked to complete the tolerance level questionnaire and the sentence scrambling task. Subsequently, the experimenter guided the participant to the VR room. There, the experimenter explained the task: the participant would have to walk through a virtual hospital room with a patient in bed. Patients had one of the two diagnoses, i.e. psychosis vs. broken leg. Participants had to say the patient's name, which was stated at the wrist of the patient, and the diagnose. It was mentioned that the name of the patient would switch in each trial. To start the next trial, the participant had to walk to a red dot in the the middle of the room.

After explanation, the experimenter helped to put on the equipment. The participant was told to first explore the empty room and then go to the red dot to start with the experiment. The experimenter observed the participant and what he or she was seeing during the experiment (i.e. on a computer screen). The experiment started with a practice trial. Here, the patient was healthy. After the practice trial, the experimenter made clear he wanted to be informed about the physical or mental condition of the patient in the beginning of each block. When the participant was

wrong with the diagnose, the experimenter explained what the diagnose of the patient was. After the experiment, the experimenter asked the participant to recall as many names as possible, this was to make sure the coverstory worked well. This surprise recall was included because otherwise, the participants could doubt the memory task they had to perform, because it was too easy (they just had to keep one name in mind for a couple of seconds). Now, they really had to recall, which made the task believable. The experiment consisted of 21 trials (i.e. one practice trial and 20 experimental trials).

### *Manipulation check and debriefing*

At the end of the experiment, the participant was asked to guess the goal of the experiment. After this, the participant was debriefed and the experimenter thanked the participant for participation in the experiment.

### **Design and statistical analysis**

In the sentence scrambling task participants were randomly assigned to one of the two conditions: the priming condition or the non-priming (control) condition. In the VR, there were two patients: one with a broken leg and one with a psychosis. The experiment consisted of 20 trials: first ten consecutive trials with the one and the same patient in all ten of these, followed by ten trials with the other patient. The participants were either in the order condition starting with the psychosis (psycho-leg), or in the order condition starting with the broken leg (leg-psycho). Participants were evenly distributed across conditions. We counterbalanced the conditions by usage of a balanced latin square. The design is a 3x3 factorial design. We used a repeated measures mixed model ANOVA. Within and between factors were distinguished. The between factors were order (psycho-leg or leg-psycho) and condition (priming or non priming). In addition to that, the diagnose (broken leg vs. psychosis) was the within subject factor. Mean minimal distance (MMD) towards the wrist in both diagnoses (i.e. broken leg vs. psychosis) in the VR environment was used as the dependent variable in the experiment.

## **RESULTS**

### **Internal validity of the questionnaire items**

At first a factor analysis was done to investigate whether the questionnaire measuring tolerance was valid. Upon inspection of the eigenvalues, the screeplot and the factorloadings in the factor matrix (i.e. table 3), and after exploration of the content of the items, we finally distinguished two factors: internally motivated tolerance items (1, 3, 4, 5, 6 and 8), and externally motivated tolerance items (2, 7, 9 and 10). Internally motivated means the motivation for the behavior described in a specific item is to follow ones own values; externally motivated means its motivation is to satisfy (the values of) others/society or to be liked by others. The factor analysis



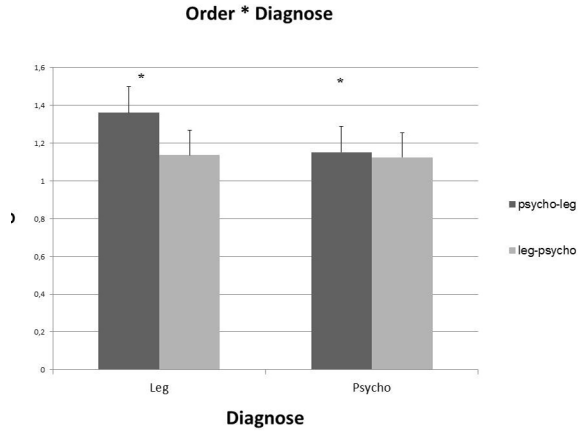
showed that item 2 was an internally motivated tolerance item, whereas the content of this item showed that this item was an externally motivated item. We decided to bring this item under the externally motivated tolerance items, because the loadings of this item did not differ extremely. The items were reliable, Cronbach's ( $\alpha = 0,809$ ) was found in the internally motivated tolerance items ( $n=6$ ) and Cronbach's ( $\alpha = 0,704$ ) was found in the externally motivated tolerance items ( $n=4$ ), see table 3. The centered internally and externally tolerance items were included as covariate in the analysis.

**Table 3:** Oblimin rotated factor matrix of the items measuring tolerance. Note: the boldface denotes the strongest factor loading for each item.

Items	IM	EM
Tolerance 1	0.777	0.134
Tolerance 2	0.526	0.352
Tolerance 3	0.754	-0.229
Tolerance 4	0.625	-0.178
Tolerance 5	0.351	-0.370
Tolerance 6	0.765	0.035
Tolerance 7	0.247	0.738
Tolerance 8	0.622	-0.119
Tolerance 9	0.027	0.935
Tolerance 10	-0.074	0.485

### Mixed Model ANOVA

More participants were in the priming condition ( $n=12$ ) than in the non-priming condition ( $n=9$ ). This was because we excluded the two participants after the data reduction and both excluded participants were in the non-priming condition. Initially, ten participants were in the psycho-leg order, whereas eleven participants were in the leg-psycho order. Three tests were relevant to test our main hypothesis. At first, the interaction of the  $3 \times 3$  mixed model repeated measures ANOVA within subjects effects was not significant ( $F(1, 15) = 0.015$ ,  $p = 0.904$ ). Second, the interaction of condition and diagnose was also not significant ( $F(1, 15) = 0.602$ ,  $p = 0.450$ ). Moreover, the test of between subjects effect of condition was not significant ( $F(1, 15) = 0.018$ ,  $p = 0.896$ ), which states that priming did not work. However, the interaction of order and diagnose was significant ( $F(1, 15) = 5.071$ ,  $p = 0.040$ ), see figure 1.



**Figure 1:** Interaction effect of diagnosis and order, with mean minimal distance towards the wrist on the vertical axis and condition on the horizontal axis (\* significant  $p < 0.05$ ).

The interaction between order and diagnosis was further explored. We decided to split the file based on order. In the psycho-leg order we found a significant effect for diagnosis ( $F(1, 6) = 12.929$ ,  $p = 0.011$ ). MMD was larger in the leg-psycho order. In the psycho-leg order a trend for interaction between internally motivated tolerance items and diagnosis ( $F(1, 6) = 2.068$ ,  $p = 0.200$ ) was found. However, in the leg-psycho order no significant effect for diagnosis was found ( $F(1, 7) = 0.193$ ,  $p = 0.647$ ), neither a trend of interaction between diagnosis and internal tolerance  $F(1, 7) = 0.005$ ,  $p = 0.946$  was found.

### Manipulation check

None of the participants guessed the goal of the current study. We therefore can conclude that the manipulation worked well.

## DISCUSSION

The goal of the present study was to test whether tolerance priming can have an influence on the approaching behavior to psychotic persons. Furthermore we wanted to investigate whether the effect of priming tolerance is dependent on the personal tolerance level.

Our data show that priming has no measurable influence on behavior in social situations. In spite of insignificance, our data still made a contribution to specify whether priming is suitable to improve approaching behavior. The data reveal important information about human approaching tendencies and provide a basis to improve further research.

The absence of a measurable influence of priming could be due to the fact that we used the wrong priming technique. Different techniques using subliminal messages or a word scramble task could have been more effective. Furthermore, it could have happened that the prime did not activate additional information in the mind's facet. We tested psychology students, who may be more tolerant towards individuals suffering from a mental disorder than other students. According to Wheeler and Berger (2007) a prime can have little influence if the trait primed already is active (Wheeler & Berger 2007). Thus, if the tolerant facet of the self already is active because psychology students naturally may be more tolerant, the prime cannot elicit more tolerance than is already active. In conclusion, the data could have been different with another sample.

All in all, participants kept more distance to the person with the broken leg. This was contrary to our expectations because we thought that the participants would keep more distance to the psychotic patient. This pattern could be the result of a diversity of reasons. First, it could have been that the participants were more cautious with regard to the person with the broken leg, because of the knee crutch. This knee crutch was standing outside of the bed and in the direction to the participants, so it could have been that they were keeping more distance trying not to bump into it.

Second, may be possible, that the participants were feeling more secure when confronted with the psychotic patient. This patient was fixed on the bed and the participants could have been sure that he could not stand up or approach them. This could result in less distance due to less fear of physical confrontation (Corrigan et al., 2001).

In addition to that, we discovered an order effect of diagnose, but only in one direction. The participants kept more distance to the patient when they first saw the psychotic patient and then the one with the broken leg. Several reasons for that pattern of results exist. First, it could be that the participants thought that the patient with the broken leg would behave differently than the fixated patient in that he might stand up. Secondly, they may have been influenced by the previous diagnose. Our participants maybe were thinking that the second patient could have been sickened with broken leg and a mental disease like the first patient.

Nevertheless, all those explanations are post hoc explanations. We have no theoretical basis for those speculations and it requires further research to test whether one of those explanations is suitable.

Several limitations are present in the current design. One limitation concerns the type of measurement we used. Only approaching behavior was measured instead of a complex interaction. The results could be different when a more complex interaction (e.g. a conversation) would be used. Unfortunately, such a complex situation is not possible within VR. In addition to that, VR has technical limitations. Sometimes the view of the participant was disturbed by lags. This might led the participant repeatedly realize that this situation was virtually. Furthermore, the participants had to wear the backpack equipped with the laptop and wires which also could have had an influence on the natural movement. Moreover, our VR did not integrate any sounds which maybe reduced the realistic experience of the situation. In addition to that, the room was not equipped as one would expect a hospital room to be. It was very empty and contained not as much equipment as a hospital room normally has.

Another limitation concerns the difference between individuals in distance. The distance our participants kept may have been influenced for two different reasons. First different approaching tendencies could be caused by gender differences. The patient lying in bed always was a man. Maybe, our results could have been different when a woman would have been lying there. According to Bailenson, Blascovich, Beall & Loomis (2001) the distance kept differs between men and women. Women tend to keep less distance to a woman, whereas men tend to keep the greatest distance to a man. Moreover, a different tendency between men and women performing the task could exist. It was experienced that women nearly always kept the same distance, whereas men kept more and more distance as the experiment progressed. This could be due to greater habituation within men. The habituation could, in fact, influence the internal validity of our distance measurement. Second, variation in distance could result from different eyesight the participants had. This could be a threat to our external validity, because every participant might have kept his/her unique distance due to different vision.

Some methodological limitations occurred during testing. First, it was not mentioned that there is a difference between the two patients lying in bed. After experiencing that our first participant did not see a difference between the two patients, we asked the following participants what they think the person was sickened with. The answers revealed that the psychotic person was not obvious enough to be directly seen as psychotic. Maybe a more extreme appearance would have been useful. Second, we did not randomize the different patients, all participants got blocks out of 10 trials with the same diagnose. It would have been better to switch the diagnoses during the task. The sequence, in which patients first occurred, was counterbalanced. Nevertheless, a randomization would have increased our internal validity. Third, no manipulation check of the priming procedure was used. This check would have been useful to test whether priming elicited more tolerance. Fourth, our power was threatened by the sample size ( $n = 21$ ), so it would be useful to test more participants to underline the results in future research.

All in all, these data revealed that priming of tolerance utilizing a sentence scramble task did not have an influence on the approaching behavior to mentally diseased patients in a VR environment. Maybe further research could reveal an effect of priming, if there is taken care of the limitations we found in our design. A different priming technique, another social situation or measurement could elicit a priming effect. In conclusion, all of our limitations should be taken into consideration in VR research as well as in priming and approaching behavior research.

Nevertheless, priming has much potential to be used as a supportive technique for improvement of behavior. It is important for human functioning to have social guidance on how to react and priming could provide some guideline (Kawakami et al., 2003). For instance, priming could encourage more social behavior in youth programs, for example anti bullying campaigns or sex education. In addition to that, priming could be used for health promotion purposes like AIDS campaigns or anti addiction programs. It requires much more research about how to apply priming and what technique is the best, but we should not underestimate its potential.

## Acknowledgements

We would like to thank our supervisor Rob Ruiter for his support and guidance during the whole research project. Furthermore, we would like to thank Henna Toppenberg for the provision of her script for VR and her assistance during analysis and other research processes. Moreover, we would like to thank Richard Benning for editing the script to fit our purposes.

## REFERENCES

- American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.
- Bailenson, J. N., Blascovic, J., Beall, A. C. & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence*, 10 (6), 583- 598.
- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71 (2), 230-244.
- Corrigan, P.W., Green, A., Lundin, R., Kubiak, M.A. & Penn, D.L. (2001). Familiarity With and Social Distance From People Who Have Serious Mental Illness, *Psychiatric Services*, 52, 7, 953-958.
- Hansen, J. & Wäncke, M. (2009). Think of capable others and you can make it! Self-efficacy mediates the effect of stereotype activation on behavior. *Social Cognition*, 27(1), 76-88.
- Howarth, P.A., & Costello, P.J. (1997). The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18, 107-116.
- Kawakami, K., Dovidio, J. F., Dijksterhuis, A. (2003). Effect of social category priming on personal attitudes. *Psychological Science*, 14 (4), 319.
- Lamy, L., Fischer-Lokou, J. & Guéguen, N. (2012). Priming emotion concepts and helping behavior: How unlivd emotions can influence action. *Social Behavior and Personality*, 40(1), 55-62.
- Lauber, C., Nordt, C., Falcato, L. & Rössler, W. (2004). Factors Influencing Social Distance Toward People with Mental Illness. *Community Mental Health Journal*, 40 (3), 265-274.
- Li, W., Moallem, I., Paller, K. A. & Gottfried, J. A. (2007). Subliminal Smells Can Guide Social Preferences. *Psychological Science*, 18 (12), 1044-1049.
- Plant, E.A., & Devine, P., (1998). Internal and External Motivation to Respond Without Prejudice. *Journal of Personality and Social Psychology*, 75 (3), 811-832.
- Piskulic, D. & Addington, J. (2011). Social cognition and negative symptoms in psychosis. *Psychiatry research*. 188, 283-285.
- Stout, P. A., Villegas, J. & Jennings, N. A. (2004). Images of Mental Illness in the Media: Identifying gaps in the Research. *Schizophrenia Bulletin*, 30(3), 543-561.
- Vaughan, G. & Hansen, C. (2004). 'Like Minds, Like Mine': a New Zealand project to counter the stigma and discrimination associated with mental illness. *Australasian Psychiatry*, 12 (2), 113-117.
- Wheeler, S. C. & Berger, J. (2007). When the Same Prime Leads to Different Effects. *Journal of Consumer Research*, 34, 357-368.
- Wheeler, S. C., DeMarree, K. G., Petty, R. E. (2007). Understanding the Role of the Self in Prime-to-Behaviour Effects: The Active-Self Account. *Personality and Social Psychology*, 11(3), 234-261.

