Determinants of event-related potentials during deception paradigms

Investigating individual differences and effects of the moral context

-Kumulative Arbeit-

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von

Vera Julia Scheuble

aus Siegburg

Bonn, 2022

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:

Prof. Dr. Rainer Banse (Vorsitzender)

Prof. Dr. André Beauducel (Betreuer und Gutachter)

Prof. Dr. Ulrich Ettinger (Gutachter)

PD Dr. Ralf Dohrenbusch (weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: Freitag, der 01. Juli 2022

Contents

Danksagung

Ich freue mich, dass ich auf meinem Weg zur Promotion nicht alleine war, sondern mich hierbei einige Personen begleitet und unterstützt haben.

Zuerst und ganz besonders gilt mein Dank meinem Betreuer *Prof. Dr. André Beauducel*. Vielen Dank, dass du mich zu Beginn meiner Promotion ermutigt hast im EEG-Labor zu arbeiten, dass du meine Fragen und Anmerkungen immer ernst genommen hast und für deine schnellen und kompetenten Antworten. Von deinem hilfreichen Feedback und insbesondere guten methodischen Ideen konnte ich viel lernen.

Weiterhin danke ich *Prof. Dr. Ulrich Ettinger* sehr dafür, dass er sich die Zeit dafür nimmt meine Arbeit zu begutachten. *Prof. Dr. Rainer Banse* und *PD Dr. Ralf Dohrenbusch* danke ich für die Bereitschaft, Teil meiner Prüfungskommission zu sein. Vielen Dank auch Ralf für ehrliche Gespräche und ermutigende Worte.

Vielen Dank an alle meine Kollegen und Kolleginnen der Abteilung Methodenlehre *Anneke Weide, Dr. Nicolas Haverkamp, Dr. Tobias Kuhl, Mathias Krüger, Christine Schoepe, Sonja Lau, Dr. Uwe Kleinemas, Dr. Norbert Hilger* und *Dr. Christopher Harms* für interessante Gespräche in der Mittagspause und darüber hinaus. Ich danke auch netten Kollegen aus Nachbarabteilungen, insbesondere *PD Dr. Michael Kavšek* für aufmunternde Gespräche, die zu einem guten Start in den Arbeitstag geführt haben.

Ein besonderer Dank gilt allen Mitgliedern unseres EEG-Labors, die mich bei meinen Dissertationsstudien unterstützt haben: *Jennifer Günther, Christina Faller, Sophie Corea, Corinna Hank, Martin Raaf, Katharina Bodenheim, Sophie Schormann, Matthias Mildenberger, Niklas Niedermeier, Sabine Sedlaczek, Leona Wahnschaffe, Riana Akhmetova, Leon Dehnhardt, Luise Thönes* und *Juliane Bartsch*. Es hat mir immer viel Freude bereitet, mit euch zusammenzuarbeiten. In diesem Zusammenhang möchte ich auch allen Teilnehmenden meiner ExPra Kurse danken. Ein herzlicher Dank gilt auch allen Probanden.

Schließlich möchte ich meinen Freunden und meiner Familie dafür danken, dass ich mich immer auf sie verlassen kann. Vielen Dank an meine Eltern, *Dieter* und *Lisa*, dafür, dass ihr mich während meines gesamten Studiums unterstützt habt und mir immer etwas Gelassenheit geschenkt habt, die gerade in stressigen Zeiten nötig war. Vielen Dank an meine Geschwister *Anne* und *Pascal* sowie meine Neffen und Nichte für gute Erholungspausen. Zuletzt danke ich *Benjamin* für unzählige Gespräche über meine

Dissertation und deine Unterstützung jeglicher Art, sei es Rat zur Programmierung, englischen Formulierungen oder dein wertvoller emotionaler Rückhalt. Vielen Dank!

Abstract

Paradigms based on event-related potentials (ERPs) are promising for detecting deception. However, less is known about moderators of the ERP patterns during deception tasks. To ensure that the investigated ERPs are markers for deception in multiple situations and for different individuals, their moderating variables need to be unraveled. Furthermore, this allows to better understand the cognitive processes involved in deception tasks. Therefore, in three studies, determinants of P300s and medial frontal negativities (MFNs) in deception tasks were investigated. A special focus lied on variables related to morality.

ERP-based deception studies typically apply the concealed information test (CIT): Participants conceal knowledge about probe items, which they have seen before, e.g., during a mock-crime, and honestly indicate for irrelevant items that they do not know them. In two of the thesis' studies, the situations before the CIT, during which participants got to know the probe items, were manipulated. In one study, the moderating effect of moral involvement was analyzed: Some participants witnessed, and others demonstrated, a behavior causing a small social problem. Additionally, I investigated whether Machiavellianism moderates the patterns of P300 and MFN amplitudes. In another study, the moderating effect of moral valence was examined. Participants saw the probe items while performing a negative or positive behavior (committing mock-theft vs. giving a present). In a third study, participants lied about their attitudes. The moderating effect of Machiavellianism on P300 and MFN amplitudes was also investigated for this deception paradigm.

Overall, MFN amplitudes were enlarged for deceptive compared to honest responses, indicating stronger response conflicts for deception. As expected, the P300 displayed a dual-nature in the two deception paradigms. For CITs, P300s were enlarged for probe items requiring a deceptive response, compared to irrelevant items requiring an honest response, revealing a greater salience of probe items. In the paradigm that did not involve the concealment of knowledge but deception about attitudes, P300 amplitudes were suppressed for deceptive compared to honest responses, revealing a greater mental workload for deception. Whereas moral valence did not moderate patterns of MFN amplitudes, a moderation effect occurred for moral involvement. The difference of MFNs between probe and irrelevant items was reduced for participants witnessing a behavior

causing a problem than for participants demonstrating this behavior. Accordingly, conflicts during deception were attenuated for witnesses. In general, Machiavellianism did not moderate the conflicts during deception, as indicated by MFN amplitudes, but did so in one condition in which the benefits of deception could be perceived as high. Patterns of P300 amplitudes proved to be stable for differences in the variables related to morality. P300 amplitudes seem to be promising markers of deception, even in social, non-forensic situations. They were unaffected by Machiavellianism, situations with a positive and negative moral valence, witnesses as well as people performing a behavior causing a small social problem.

Keywords: deception; CIT; MFN; P300; Machiavellianism; morality

Zusammenfassung

Ereigniskorrelierte Potenziale (EKPs) gelten als erfolgversprechend für die Aufdeckung von Täuschung. Es ist aber noch wenig darüber bekannt, von welchen Variablen die EKPs während Täuschungsaufgaben beeinflusst werden. Um sicherzustellen, dass mittels EKPs Täuschung in unterschiedlichen Situationen und für verschiedene Personen erkennbar ist, müssen deren Einflussgrößen aufgedeckt werden. Dies bietet gleichzeitig die Möglichkeit, ein besseres Verständnis von den während Täuschung ablaufenden kognitiven Prozessen zu erlangen. In drei Studien wurden die Muster der P300 und medialen frontalen Negativität (MFN) sowie deren Moderatorvariablen in Täuschungsparadigmen untersucht. Ein besonderer Fokus lag auf dem moralischen Kontext der Täuschungsaufgabe.

In EKP-basierten Täuschungsstudien wird meist der Tatwissenstest eingesetzt. Die Probanden verheimlichen, dass sie bestimmte Items kennen (sogenannte Probe-Items), welche sie zuvor gesehen haben, beispielsweise während eines gestellten Verbrechens. Bei anderen, irrelevanten Items geben sie dagegen ehrlich an, dass ihnen ebendiese unbekannt sind. In zwei Studien der Dissertation wurde die Situation vor dem Tatwissenstest verändert, während der die Probanden den Probe-Stimulus kennenlernen. In einer Studie wurde die moralische Involviertheit der Probanden manipuliert: Einige Probanden übten ein Verhalten aus, das zu einem kleinen sozialen Konflikt führte. Dagegen beobachteten andere Probanden eine Person, welche dasselbe Verhalten zeigte. Zusätzlich wurde analysiert, ob Machiavellismus die Muster der P300- und MFN-Amplituden beeinflusst. Innerhalb einer weiteren Studie wurde der Effekt von moralischer Valenz auf die beiden EKPs untersucht. Die Probanden sahen den Probe-Stimulus entweder während sie ein negativ oder positiv konnotiertes Verhalten ausübten (einen Diebstahl verüben vs. einer anderen Person ein Geschenk geben). In einer dritten Studie logen die Probanden über ihre eigenen Einstellungen. Auch für dieses Täuschungsparadigma wurde analysiert, ob Machiavellismus die P300- und MFN-Amplituden beeinflusst.

In allen drei Studien ging Lügen mit negativeren MFN-Amplituden einher als ehrlich zu antworten, was darauf hinweist, dass Täuschung mit stärkeren Antwortkonflikten verbunden war. Wie erwartet zeigten sich unterschiedliche Muster der P300-Amplituden in den beiden Täuschungsparadigmen. Beim Tatwissenstest fanden sich positivere P300-Amplituden für Probe-Items im Vergleich zu irrelevanten Items. Probe-Items waren demzufolge salienter. Wenn es dagegen nicht um die Bekanntheit von Stimuli ging, sondern

die Probanden über ihre eigenen Einstellungen logen, zeigten sich verminderte P300-Amplituden für Lügen im Vergleich zu ehrlichen Antworten. Dies weist darauf hin, dass Lügen mit einer erhöhten kognitiven Anstrengung verbunden war. Für Zeugen eines Verhaltens, das zu einem kleinen sozialen Konflikt führte, zeigte sich ein geringerer Unterschied der MFN-Amplituden zwischen irrelevanten und Probe-Stimuli als für Probanden, die das Verhalten selber ausübten. Für Zeugen war Lügen dementsprechend mit geringeren Antwortkonflikten verbunden. Es fand sich kein allgemeiner Einfluss von Machiavellismus auf die MFN-Amplituden, sondern nur in einer Bedingung, in der die Vorteile der Täuschung als hoch eingeschätzt werden konnten. P300-Amplituden wurden nicht von den untersuchten moralischen Variablen beeinflusst. Daher scheinen P300-Amplituden auch in nicht-forensischen, sozialen Situationen erfolgversprechend für die Aufdeckung von Täuschung zu sein. Sie wurden nicht von der Machiavellismusausprägung der Probanden beeinflusst, ebenso nicht von der moralischen Valenz der Situation sowie der Involviertheit der Probanden bei einem Verhalten, das einen kleinen sozialen Konflikt verursachte.

Schlagwörter: Täuschung; Tatwissenstest; MFN; P300; Machiavellismus; Moral

1 Introduction

"Lies, my dear boy, can easily be recognized. There are two kind of them: those with short legs, and those with long noses. Your kind have long noses."

—Carlo Collodi, Pinocchio

"Rather than love, than money, than fame, give me truth."

—Thoreau, Walden, 1854

Deception is common in everyday life (Vrij, 2008). On average, people tell 1-2 lies[1] per day and lie in every fourth social interaction (Hancock et al., 2004; Kashy & DePaulo, 1996; Serota et al., 2010). As frequent as lies are, as various are also the situations in which people lie. Pretending to like a distasteful homemade meal, acting delighted when faced with an unpleasant present, or covering up a surprise for a friend are just some examples in which lies could be considered as harmless or even be told with the intent to prevent harm from others. But there are also lies that are mainly told for egoistic reasons. For instance, to cover up one's own misbehavior. This seems especially apparent in the forensic context when people try to cover up a theft or a violent crime. Yet, even in the forensic context, situations in which people lie are manifold, and some lies are also told to benefit others. This can be the case for sham marriages (D'Aoust, 2013), animal right extremists freeing animals (J. V. Carson et al., 2012), or physicians knowingly filling out their patients' clinical reports incorrectly to provide them with the care they consider necessary (Wynia et al., 2000).

The fact that people can lie in every social interaction makes them more complex for both the liar and the recipient of the lie. When deciding to speak the truth, a person can respond to an answer with the words coming directly into mind. When deciding to lie, the truth has to be inhibited and a different response has to be given, which is therefore cognitively more challenging (Walczyk et al., 2003). Lies are usually seen as morally reprehensible, especially when they are told for selfish reasons (Backbier et al., 1997; Lavoie et al., 2016; Robinson, 1994). Accordingly, liars possibly have to cope with feelings of remorse and moral conflicts. The fact that people lie frequently also means that

[1] In line with the wording in previous deception literature (DePaulo et al., 2003; Ekman; 2001; Johnson et al., 2008; Rosenfeld, 2020; Rosenfeld et al., 2012; Suchotzki et al., 2015; Vrij, 2008), I use the terms *lying* and *deception* interchangeably in the thesis.

we cannot trust in every word another person says. There is always the possibility that a lie was told. Especially when personal involvement in the situation is high, a careful check whether the other person is saying the truth and the told story is plausible seems to be needed. How much easier life could be in such situations if one could detect a lie as in the tale of Pinocchio by the other person's tip of the nose or as in the wizarding world of Harry Potter where some drops of a truth serum could ensure that the truth is being told.

1.1 Deception detection methods: a brief overview

Throughout history people have been fascinated by deception and its possible detection (Ford, 2006; Trovillo, 1939). Investigating cognitive, physiological and emotional processes during deception went hand in hand with finding ways to detect deception. Already deception detection methods dated back as early as about 1000 B.C. relied on the assumption that deception reflects itself in the bodily activity: In China people suspected to be lying had to chew rice powder and subsequently spit it out (Kleinmuntz & Szucko, 1984; Trovillo, 1939; Vrij, 2008). It was assumed that fear is accompanied by a dry mouth (Ford, 2006). Accordingly, if the rice powder remained dry, this was seen as a sign that the person was lying (Ford, 2006). Since then, deception has been further investigated and various methods for deception detection have been proposed. They can be broadly categorized into deception detection based on non-verbal, verbal, physiological and neural measures (Granhag et al., 2015; Granhag & Strömwall, 2004).

When people have to judge whether others are lying, they are only slightly better than chance in detecting deception (Bond & DePaulo, 2006; Vrij, 2008). Also, experts in deception (i.e., people who have to deal with deception regularly at work, like police officers or judges) do not significantly outperform laypeople (Bond & DePaulo, 2006). To improve deception detection based on non-verbal behavior, Ekman and colleagues proposed to analyze facial expressions of emotions in detail. More precisely, one should pay attention to facial expressions with a very short duration, so called microexpressions (Ekman, 2003; Ekman & Friesen, 1969). According to Ekman, emotions are accompanied by specific facial muscle activities (Ekman, 2003; Ekman & Friesen, 1969). When people try to cover up their real emotions, it is possible that they involuntarily leak their real emotions via their facial expression for a very short time before they manage to mask them (Ekman, 2003; Ekman & Friesen, 1969). Yet, not all people display

microexpressions and even though microexpression can possibly give hints for the real emotion of a person, they still do not reveal why a person feels this emotion, which needs to be unraveled to know the exact content of the lie (Ekman, 2009).

A prominent method to assess the veracity of a statement by its verbal content is *statement validity assessment* (SVA). SVA was originally applied for testimonies of children about sexual abuse (Köhnken, 2004; Vrij, 2008). It consists of several steps: Based on a case-file analysis a semi-structured interview is conducted (Köhnken, 2004; Vrij, 2008). The core component of SVA is the criteria-based content analysis: The credibility of the transcribed statements of the interview is assessed by 19 criteria. They involve, for instance, whether the said was logically consistent and included many details (Granhag & Vrij, 2005; Köhnken, 2004; Vrij, 2008). Finally, alternative explanations for the fulfillment or absence of the criteria in the statements are checked, e.g., influence of other persons and preparation for the interview (Vrij, 2008). It has to be noted that the outcome of an SVA also relies on the expertise of the interviewer and the evaluator of the transcribed interview (Brigham, 1999; Lamb et al., 2008).

Physiological and neural measures come with the promise to provide a more objective method to detect deception, especially since they are often combined with a computerized questioning of examinees. A well-known physiological method is the polygraph. It is also known as “lie detector”. However, a word of caution is needed for this naming: The polygraph does not directly detect lies per se. Like all outlined methods intended to detect deception, it indicates processes accompanied by deception (Vrij, 2008). The polygraph measures bodily activity, like changes in blood pressure, respiration, and electrodermal activity (Ben-Shakhar & Elaad, 2003; Granhag & Vrij, 2005). A relatively well-researched and recommended test to be applied with the polygraph is the concealed information test (CIT; Ben-Shakhar & Elaad, 2003; Iacono, 2014; Meijer & Verschuere, 2015). As its name suggests, the CIT serves to reveal concealed knowledge. In the CIT, irrelevant unknown items are shown together with a probe item that is meaningful and known only to certain people. For people who recognize the probe item and have to conceal their knowledge about it during the CIT, the probe stands out from the other irrelevant items. This goes along with different physiological responses for probe and irrelevant items (Ben-Shakhar, 2012; Meijer et al., 2007; Rosenfeld et al., 1991). For

instance, thieves could therefore be uncovered by their recognition of objects from the crime place.

1.2 The focus of the present thesis

Recently, the relevance and potential of neural measures to detect deception has been highlighted (Ford, 2006; Ganis & Keenan, 2009). In the last decades, an increasing number of ERP and functional magnetic resonance imaging (fMRI) studies emerged, investigating cognitive processes involved in deception and ways to detect deception. Meijer et al. (2014) found in their meta-analysis that an ERP that is regularly analyzed in deception studies, the P300 amplitude, is as good as polygraphs in revealing concealed knowledge and even outperforms them for some item types. As a result, some researchers suggest that, in the future, ERP amplitudes should be used for detecting deception in practice (Farwell et al., 2013; Johnson, 2014; Meixner, 2018). However, for their application, it seems crucial to know which variables affect ERP amplitudes in deception tasks. Possible confounding variables have to be unraveled in order to minimize the risk of errors. This line of research also offers the possibility to gain a better understanding of the cognitive processes involved in deception.

The focus of the thesis' studies lies therefore on ERP patterns accompanied by deception and their determinants. As in the quote from Pinocchio, we investigate two possible markers of deception: the P300 amplitude and medial frontal negativity (MFN). It should be noted that ERPs are not direct indicators of deception, like the long noses and short legs in Pinocchio. They reflect cognitive processes that accompany deception. The P300 amplitude indicates, depending on the deception paradigm, an increased salience of concealed knowledge about an item or a greater mental workload for deceptive compared to honest responses (Johnson et al., 2008; Leue et al., 2012; Vendemia & Buzan, 2005). The MFN indicates conflicts, which are typically increased during deceptive compared to honest responses (Johnson et al., 2008; Leue et al., 2012). As the P300 amplitude and the MFN are both related to deceptive responses in the CIT, it should be investigated whether these ERPs are similar in different CIT contexts or whether they vary for different individuals and situations. As outlined in the beginning of the introduction, situations in which people lie are manifold and, in some instances, lies can be perceived as more or less acceptable. In other words, situations in which people lie vary in their moral

reprehensibility. Hence, a central aim of the thesis was to elucidate whether MFN and P300 patterns differ depending on the moral context of the deception task.

But not only situations in which people lie vary. People also differ in their perception of lies. Some people have less scruple to lie than others. Machiavellianism is a trait that is closely associated with both a tendency to be deceptive and a disregard of conventional morality (Christie & Geis, 1970; Jones & Paulhus, 2009). It might therefore be expected that individuals differing in Machiavellianism also differ in the cognitive processes accompanied by deception. If an item that has to be concealed is less salient or deception is accompanied by fewer conflicts for individuals higher in Machiavellianism, this could result in attenuated P300 and MFN amplitudes, respectively. In this case, it would be more difficult to differentiate between deceptive and honest responses based on P300 and MFN amplitudes for individuals higher in Machiavellianism. This seems highly relevant for the application of ERPs in deception detection. Hence, a further aim of the thesis was to investigate whether Machiavellianism moderates P300 and MFN amplitudes during deception tasks.

1.3 Structure of the thesis

When investigating ERP patterns during deception, it first seems essential to know which behaviors fall under the term deception. Accordingly, the theory section starts with an outlining of defining characteristics of deception (Chapter 2.1). Since ERPs elucidate cognitive processes and the present thesis focuses on cognitive processes during deception tasks, the following chapter deals with research findings and theories of cognitive processes involved in deception (Chapter 2.2). Subsequently, Chapter 2.3 provides an overview of the EEG, ERPs in general, and the ERPs investigated in the thesis, i.e., the P300 and MFN, and their role in deception paradigms. The most prominent and probably also a promising paradigm for the detection of deception based on ERPs is the CIT (Iacono, 2014; Rosenfeld et al., 2013). Therefore, Chapter 2.4 deals with the CIT and its guiding principles. In all studies of the thesis, moderating effects of variables related to the moral context of the deception paradigm are investigated. Hence, Chapter 2.5 gives a brief overview of moral theories, morality in the context of deception and the personality trait in the focus of the thesis, Machiavellianism. This leads to the outlining of the

overarching goals and research questions of the studies and a description of methods central to all studies of the thesis (Chapter 2.6 and 3).

The studies of the thesis are presented in Chapter 4. In all studies, P300 and MFN amplitudes are investigated during deception paradigms. Study I focuses on individual differences in patterns of P300 and MFN amplitudes during a CIT by analyzing the moderating effect of moral involvement (witnessing vs. demonstrating a behavior leading to a social conflict) and Machiavellianism. In study II, P300 and MFN amplitudes and moderating effects of Machiavellianism are investigated for a deception paradigm that does not involve the recognition of items but deception about attitudes. Finally, study III focuses on the moderating effect of moral valence on P300 and MFN amplitudes during a CIT (giving a present vs. stealing an object and concealing knowledge about it). In the last chapter of the thesis, the results of the studies are interpreted as a whole and integrated into the existing literature. Their limitations are outlined and outlooks for future studies are derived.

2 Theoretical background

2.1 Defining deception

At first thought, the idea may arise that deception means saying something that is untrue. Yet, when defining deception by the actual falsehood of a statement, mistakenly saying something wrong as well as irony would be classified as deception. Accordingly, this definition does not seem to catch the gist of deception. Indeed, many definitions in the literature highlight that deception is an intentional act to mislead others (DePaulo et al., 2003; Ekman, 2001; St Augustine, 395/1995). Already in early discourses about deception, the perspective of the deceiver was considered: For instance, St Augustine (395/1995) described that deceivers have something in their mind they believe to be true, yet they express–either verbally or non-verbally–something else. Vrij (2008) proposed a definition of deception, which incorporates this concept and specifies important additional aspects of deception. Vrij (2008, p. 15) defines deception as "a successful or unsuccessful deliberate attempt, without forewarning, to create in another, a belief which the communicator considers to be untrue." This definition does not exclude non-verbal behavior. A deceptive response does not have to be stated. Both verbal and non-verbal behavior can be deceptive. Likewise, intentionally hiding information counts as deception (Vrij, 2008). Furthermore, it is not decisive whether the communicator is successful. Even the attempt to mislead others is considered deceptive (Vrij, 2008). Ekman (2001) highlighted that it is important that the behavior of actors and magicians is not subsumed under deceptive behavior. This aspect is also included in the outlined definition by specifying that the attempt to mislead another is given without forewarning. Vrij (2008) deliberately did not state that deception harms the receiver, since some lies can be told with the intent to benefit others. Deception is also not restricted to deceptive behavior of humans (Vrij, 2008). There are some examples in which behavior of animals or even plants can be categorized as deceptive. For instance, when observing chimpanzees during feedings, it has been found that a female chimpanzee demonstrated deceptive behavior to hide some leftover food from the other chimpanzees, which they had missed (DeWaal, 1986). Yet, the present thesis focuses only on deceptive behavior of humans. Some researchers, for example Bok (1978, p. 13), define lying as a subcategory of deception, by outlining that

lying is "any intentionally deceptive message which is *stated*." However, in the deception literature on which the present thesis is based upon, the terms *deception* and *lie* are mainly used interchangeably (DePaulo et al., 2003; Rosenfeld, 2020; Rosenfeld et al., 2012; Suchotzki et al., 2015; Vrij, 2008). Therefore, I also follow this approach.

2.2 Deception from a cognitive perspective

The notion of St Augustine (395/1995) that deception requires keeping in mind two views already points out that deception is cognitively demanding. When a question is asked, the truth comes into mind (Vrij, 2008). For lying, the prepotent honest response has to be inhibited (Sip et al., 2008; Walczyk et al., 2003). A liar has to come up with a new, consistent answer and remember this answer (Sip et al., 2008; Walczyk et al., 2003). During these steps, conflicts between the truth and lie have to be managed (Vrij, 2008; Johnson, 2014). Furthermore, lying sometimes requires adapting emotional expressions and handling feelings of guilt (Ekman, 2001). Verbal and non-verbal behavior contradictory to the lie has to be avoided (Suchotzki et al., 2017). In this line, working memory and especially executive processes have been associated with deception (Christ et al., 2009; Gombos, 2006; Johnson, 2014; Sip et al., 2008).

Baddeley (1992, 2010) introduced working memory as a brain system required during complex cognitive tasks to transitory store and manipulate information. Working memory consists of different components: The phonological loop and the visuospatial sketch pad store temporary memories (Baddeley, 1992). The phonological loop captures verbal information, whereas the visuospatial sketch pad captures visual images. Later, Baddeley added another memory system, the episodic buffer (Baddeley, 2010). The episodic buffer stores multidimensional episodes and can combine visual and auditory information as well as information about smells and taste (Baddeley, 2010). Furthermore, Baddeley introduced a controlling instance, the central executive (Baddeley, 1992, 2010). The central executive coordinates information of the memory systems (Baddeley, 1992, 2010). Likewise, it regulates the focus of attention and can activate information of long-term memory (Baddeley & Logie, 2016). In sum, it controls and regulates cognitive processes and is therefore in charge of executive functions (Baddeley, 1992; Miyake et al., 2000).

Miyake et al. (2000) described monitoring and updating of working memory

representations, inhibition of prepotent responses, and mental set shifting to be central executive functions. All of these processes are probably needed for deception (Christ et al., 2009): When the truth is kept in mind and a deceptive response is conceived, working memory representations have to be monitored and updated (Christ et al., 2009). The prepotent truthful answer has to be inhibited when responding deceptively (Christ et al., 2009; Johnson et al., 2008). Furthermore, mental set shifting is required when switching between deceptive and truthful responses during an interaction (Christ et al., 2009).

In a nutshell, deception is cognitively more challenging than speaking the truth. It relies on additional cognitive processes, some of which are needed to monitor and handle the additional conflicts that are prepotent when responding deceptively.

2.3 The EEG and ERPs

Cognitive processes that are prepotent during the completion of deception tasks can be indicated by ERPs. Moreover, the different cognitive processes during deceptive and honest responses can help to reveal deception through their accompanied ERP patterns. Therefore, the following chapter introduces the EEG and in a second step the ERP technique. Furthermore, two prominent ERPs in the context of deception, the MFN and P300, will be described.

2.3.1 The neural basis of EEG

The electroencephalogram (EEG) records electrical activity of the human brain (Luck, 2014; Speckmann et al., 2012). More precisely, electrodes placed on the scalp measure changes of voltage (i.e., electrical potential differences) over time (Luck, 2014). Apart from a few exceptions, the EEG reflects potential differences generated by postsynaptic potentials (see Luck, 2014; Yamada & Meng, 2018; or Zschocke, 2002, for a thorough description). The human brain consists of neurons, connected by synapses. The transmission of electrical signals between two cells takes place at the synapse (Yamada & Meng, 2018). When a neurotransmitter binds to a receptor of the membrane of a postsynapse, ion channels open or close, causing a change of the number of positively and negatively charged ions inside and outside the synapse (Luck, 2014). The electrical charges around and in the postsynapse form a dipole, that is, a separated positive and negative charge

(Luck, 2014). The potential difference between the subsynaptic and postsynaptic membrane is the postsynaptic potential (Zschocke, 2002). The voltage of a single postsynaptic potential is too small to be measurable at the scalp (Luck, 2014). Yet, the postsynaptic potentials of neurons can summate when they are active at the same time (Luck, 2014). The vector sum of the postsynaptic potential can be viewed as a single equivalent current dipole (Luck, 2014). For a measurable EEG signal, many postsynaptic potentials have to arise simultaneously at similarly oriented neurons (Luck, 2014). Due to the similar orientation of the dipoles at pyramidal cells and since their synapses often fire at the same time, their dipoles summate and their signal travels to the scalp nearly at the speed of light (Foldvary-Schaefer & Grigg-Damberger, 2012; Luck, 2014). Herein, an important feature of the EEG is already named: its high temporal resolution. The EEG records electrical activity of the scalp in the range of milliseconds (Seifert, 2005). As a reference point, the temporal resolution of imaging techniques, like fMRI and positron emission tomography (PET), is in the range of seconds (Seifert, 2005). Therefore, the EEG is also able to record very short-lasting cognitive processes of the brain (Birbaumer & Schmidt, 2010). However, the EEG is restricted to a low spatial resolution due to the fact that only the aggregated signal of many neurons is detectable at the scalp (Pape, 2009).

2.3.2 Event-related potential components

ERPs are changes of voltage recorded at the scalp related to a specific event (Luck, 2014). Moreover, ERP components are changes of voltage related to a specific event that indicate a certain neural or psychological process (Kappenman & Luck, 2010). They are systematic and reliable voltage changes originating from a single neuronal generator (Luck, 2014). This means that as long as the same neural or psychological processes occur under similar conditions, a similar change of voltage should be observed at the scalp. When conditions are altered affecting the neural or psychological process in focus, this should lead to a variation in the change of voltage. By specifying that these voltage changes stem from a single neuronal generator, it is meant that they are represented by a single dipole (Luck, 2014). This does not imply that the exact part of the brain, from which the ERP components originate, has to be known (cf. Luck, 2014). In comparison to voltage changes in the raw EEG signal, ERP components are rather small, varying around 1-30 µV (Sanei & Chambers, 2007). When collecting many epochs, meaning segments of the

continuous EEG time-locked to the onset of the event of interest, and averaging them, the unsystematically potential differences that are not related to the processing of the stimulus can be averaged out (Luck, 2014; Peng, 2019). Conversely, brain activity that is consistently related to the processing of the event remains preserved (Luck, 2014; Sanei & Chambers, 2007). Therefore, this procedure allows uncovering even small ERP components (Luck, 2014; Sanei & Chambers, 2007).

ERP components are usually characterized by their amplitude (magnitude of the neural activity), latency (timing of the amplitude/neural activation), and scalp distribution (Luck, 2005; Sanei & Chambers, 2007). ERP components that occur very shortly after the presentation of a stimulus are often exogenous components, meaning that they are primarily determined by the physical properties of their eliciting external stimulus (Fabiani et al., 2017; Birbaumer & Schmidt, 2010). Endogenous components tend to occur later and are dependent on psychological processes. They are related to processes inside the person and the interaction of the person with an event (Birbaumer & Schmidt, 2010; Donchin, 1978; Fabiani et al., 2017). Furthermore, there are also ERP components associated with the execution or preparation of a motor response (Luck, 2014). The distinction between these categories of ERP components is not always clear-cut (Luck, 2014). Some ERP components share features of more than one of these categories (Luck, 2014). For instance, the N100 is dependent on both physical properties of the stimulus and cognitive processes (Fabiani et al., 2017).

2.3.3 Why are ERP components suited to study deception?

There are several features of ERPs that are beneficial for studying deception but also some limiting factors. In the following, an overview of them will be given. Due to its high temporal resolution, the EEG enables a continuous measure of cognitive processes occurring after stimuli and responses (Johnson, 2014; Luck, 2014). Accordingly, ERPs can provide a direct measure of the neural activity between a stimulus and a deceptive response (Johnson, 2014). Even rapid sequences of cognitive processes during deception can be revealed (Luck, 2014). ERPs represent a covert measurement of processing (Luck, 2014), i.e., they do not rely on the ability of people to memorize and articulate their cognitive processes (Luck, 2014). This feature is decisive when thinking about the possibility to detect deception based on ERP components. Furthermore, EEG measures

are non-invasive and relatively inexpensive. It is therefore possible to take EEGs of many people. There are also some shortcomings, which should be considered when analyzing deception based on ERP components: ERP components rely on the averaging of many trials of the event of interest, meaning that examinees have to respond deceptively many times, which can possibly lead to an artificial situation (Luck, 2014). Furthermore, muscle movements can cause artifacts in the EEG and examinees need to be able to stand still for the time of interest (Luck, 2014).

Nevertheless, ERP studies can lead to valuable insights to the cognitive processes involved in deception. They provide complementary information to deception studies based on other measures, like response times and fMRI studies. Based on response times, inferences can be drawn about the needed time for responding deceptively, in comparison to responding honestly, and whether deception is overall cognitively more challenging than being honest (Suchotzki et al., 2017). fMRI studies are preferred when studying which brain areas are active during deception (Luck, 2005). Conversely, ERP studies are especially suited to study multiple cognitive processes involved in deception and their timing. Based on the different cognitive processes of deceptive vs. honest responses, deception can possibly be revealed. The P300 and MFN are two ERP components that have been associated with cognitive processes involved in deception tasks and will therefore be introduced in the following (Johnson, 2014; Leue & Beauducel, 2019; Rosenfeld et al., 2013).

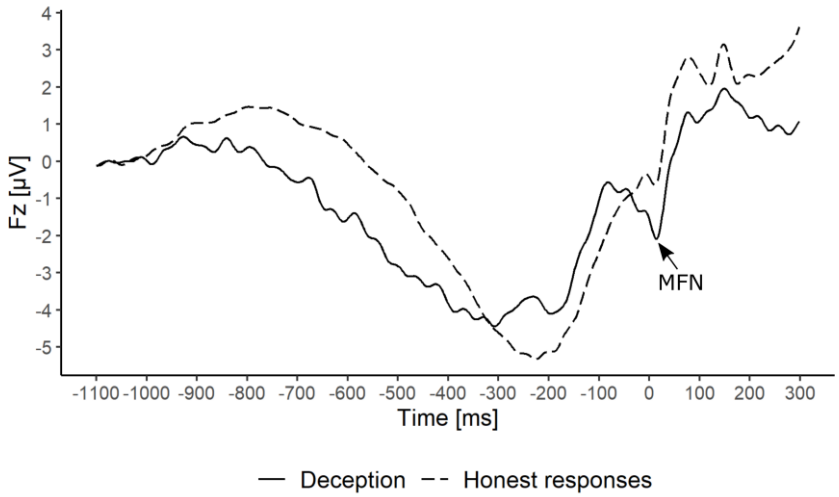
2.3.4 MFN components

MFNs are negative deflections of the ERP occurring at fronto-central electrodes (see Figure 1). They have been originally found for erroneous responses and were accordingly named error-related negativity or error negativity (ERN-Ne; Falkenstein et al., 1991; Gehring et al., 1993; Johnson et al., 2004). Yet, following studies discovered a negative deflection equivalent to the ERN-Ne for correct responses (Bartholow et al., 2005; Gehring & Knight, 2000; Vidal et al., 2000). This negative deflection, which is typically smaller than the ERN-Ne, occurred especially for trials implying a coactivation of different responses and response conflicts (Bartholow et al., 2005; Coles et al., 2001; Vidal et al., 2000). Gehring and Willoughby (2004) applied the general term *medial frontal negativities* (MFNs) for ERN and ERN-like negativities with a medial frontal scalp

distribution. Deception researchers, like Johnson and colleagues, adopted the term MFN to refer to a negative deflection at fronto-central electrodes elicited around 0-100 ms following correct responses in deception paradigms[2] (Dong et al., 2010; Johnson et al., 2004, 2005, 2008; Leue et al., 2012). It should be emphasized that in this context responding correctly is meant as responding as intended, deceptively or truthful depending on the stimulus or task block.

Figure 1

Response-locked grand averages depicting the MFN in a deception task



Note. Grand averages are based on the study by Scheuble and Beauducel (2020b), which involved a CIT.

It is suspected that the MFN is generated in the anterior cingulate cortex (ACC) itself or areas nearby (Gehring & Willoughby, 2004; Johnson et al., 2008; Nieuwenhuis et al., 2004). The MFN as well as activity of the ACC have been associated with conflicting response tendencies and response monitoring (Aron et al., 2004; Botvinick et al., 2001; Carter et al., 1998; Johnson, 2014). More generally, the ACC seems to be central for the executive control of cognition (Carter et al., 1998; Posner & Deheane, 1994; Shenhav et al., 2016). Neuroimaging studies repeatedly found an increased activation of the ACC during deceptive responses (Langleben et al., 2002, 2005; Sip et al., 2008). Likewise, ERP studies found larger (more negative) MFN amplitudes for deceptive compared

[2] Some deception researchers use the term *correct response negativity* referring to the same ERP (e.g., Suchotzki et al., 2015).

to truthful answers, revealing a greater response conflict for deception (Gibbons et al., 2018; Johnson et al., 2005, 2008; Leue et al., 2012). This is in line with the notion that when responding deceptively, conflicts are prevalent and an inhibition of the truthful response is needed (see Chapter 2.2.).

2.3.5 P300 component

The P300 component is a positive deflection usually occurring around 300-600 ms after the stimulus at parietal-central electrodes (Polich & Kok, 1995). A depiction of the P300 in a deception paradigm is given in Figure 2. P300 components are prototypically analyzed in oddball paradigms (Hruby & Marsalek, 2003). During a classic oddball task, an infrequent stimulus is presented together with other frequent stimuli (Fabiani et al., 1986; Hruby & Marsalek, 2003). The participants react–either physically or mentally–to the infrequent stimuli (Polich & Criado, 2006). The oddball (i.e., the infrequent stimuli) elicits a P300 (Hruby & Marsalek, 2003; Polich & Criado, 2006). The P300 component has therefore been associated with attentional allocation (Polich & Kok, 1995; Pritchard, 1981). Enlarged P300 amplitudes occur for salient or meaningful stimuli (Johnson, 1986). Moreover, the P300 component is sensitive to mental workload: Decreased P300 amplitudes can be found when tasks become increasingly difficult, such as in dual-tasks (Beauducel et al., 2006; M. W. Miller et al., 2011; Palmer et al., 1994).

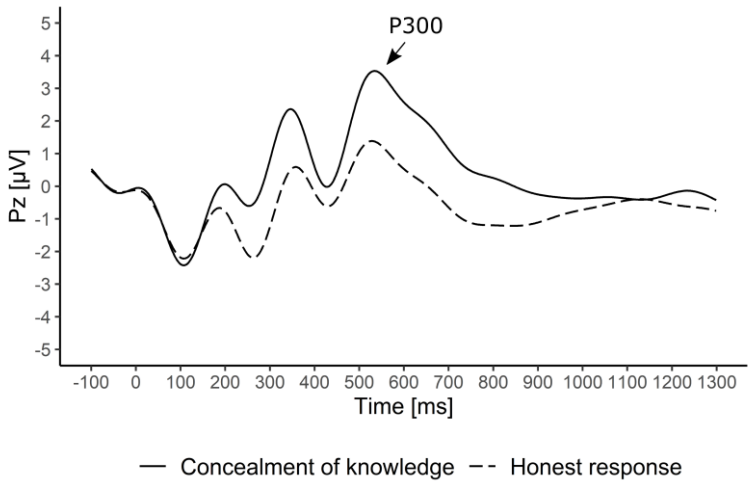
In Johnson's triarchic model (1986, 1988), different variables influencing P300 amplitudes have been grouped into three categories: subjective probability, stimulus meaning, and information transmission. Subjective probability denotes the feature that unexpected stimuli, like stimuli perceived as infrequent, are accompanied by enlarged P300 amplitudes (Johnson, 1986, 1988). For multiple stimuli forming different categories, the frequency of each stimulus category rather than the frequency of each individual stimulus is decisive for the pattern of P300 amplitudes (Johnson, 1986). Accordingly, subjective probability refers to the perceived probabilities of stimuli (Johnson, 1986). The second category, stimulus meaning, subsumes variables influencing the extent to which a stimulus is processed (Johnson, 1986, 1988). For instance, stimuli with higher personal value, such as stimuli with a higher monetary payoff, are accompanied by enlarged P300 amplitudes (Johnson, 1986). The impact of subjective probability and stimulus meaning on the P300 amplitudes is dependent on how much of the transmitted

information of a stimulus is received by a person (Johnson, 1986, 1988). When a person is distracted by an additional task and therefore does not attend to the stimulus in focus, P300 amplitudes are typically decreased (Johnson, 1986, 1988).

Most deception studies rely on the feature of the P300 to be sensitive to infrequent and meaningful stimuli (Leue & Beauducel, 2019; Rosenfeld et al., 2013). When a person recognizes a meaningful stimulus presented in a series of multiple, unknown stimuli, the known stimulus appears more infrequent and salient than the unknown stimuli (klein Selle et al., 2021; Leue & Beauducel, 2019; Meijer et al., 2007; Rosenfeld, 2011). As a result, enlarged P300 amplitudes occur for the known compared to the unknown stimuli (Leue & Beauducel, 2019; Rosenfeld et al., 2013). This pattern of P300 amplitudes also appears when the recognition of the stimulus is denied, opening the possibility to detect concealed knowledge, for instance of perpetrators of a crime (Leue & Beauducel, 2019; Rosenfeld et al., 2013). It has been argued that enhancing the awareness of deception can even increase the salience of the known stimuli leading to increased P300 amplitudes for concealed known stimuli (Rosenfeld et al., 2012, 2013; Rosenfeld, Ozsan, et al., 2017; Verschuere et al., 2009). However, it has to be noted that the results of these studies are restricted to concealed knowledge: The different patterns of P300 amplitudes are based on the concealment of the recognition of items that appear more infrequent and meaningful than other unknown items.

Figure 2

Stimulus-locked grand averages depicting the P300 in a CIT



Note. Grand averages are based on the study by Scheuble et al. (2021).

Conversely, studies analyzing P300 amplitudes in deception tasks that are not based on the recognition of an infrequent stimulus presented in a series of unknown stimuli found suppressed P300 amplitudes for deceptive compared to honest responses (Dong et al., 2010; Johnson et al., 2008; Meek et al., 2013; Pfister et al., 2014). This pattern of P300 amplitudes has been associated with greater mental workload and additional cognitive effort that is required for deceptive compared to honest responses (Johnson et al., 2003, 2005; Meek et al., 2013; Vendemia & Buzan, 2005). Deception can be considered an additional task driving the attention away from the stimulus and hampering the processing of the stimulus information, which is accompanied by suppressed P300 amplitudes (Beauducel et al., 2006; Johnson, 1986; Johnson et al., 2003, 2005; Palmer et al., 1994).

2.4 The concealed information test

The CIT is commonly applied when analyzing ERPs in the context of deception (Leue & Beauducel, 2019). It was initially named *guilty knowledge test* but is nowadays usually referred to as CIT (Lykken, 1959). The CIT serves to detect concealed knowledge (Ben-Shakhar, 2012). For instance, a murderer knows and probably tries to conceal his recognition of the murder weapon. Based on the different neurophysiological patterns of known compared to unknown items, concealed knowledge can be inferred (Lykken, 1959). In the CIT, three stimulus categories are presented, so-called probe, target and irrelevant items. The probe item is the stimulus of interest (e.g., the murder weapon). It is known only to certain people (e.g., the murderer), who deny their knowledge of the probe stimuli. Irrelevant items are similar stimuli that are unknown or without special meaning. Target items ensure that participants pay attention to the CIT by demanding a special response. During the CIT, probe items are usually presented less frequently than irrelevant items (Iacono, 2014). For people who know the probe item, it therefore reflects an oddball. Conversely, for participants who do not know the probe item, it belongs to the same category as the other irrelevant items and accordingly does not represent an oddball. The difference between known probe and unknown irrelevant items is often labeled CIT effect (klein Selle et al., 2016; Meijer et al., 2016; Rosenfeld et al., 2012).

Previous studies found enlarged P300 amplitudes for probe compared to irrelevant

items for people who know the probe items (Leue & Beauducel, 2019; Rosenfeld et al., 2013). This pattern of P300 amplitudes has been associated with a greater salience of the known probe compared to the unknown irrelevant items (klein Selle et al., 2021; Leue & Beauducel, 2019). When people recognize the probe stimulus, it is distinct and stands out from the other irrelevant items (Iacono, 2014). Note that the determinants of P300 amplitudes outlined in Johnson's triarchic model (1985) fit to the probe item: For participants recognizing the probe item, it is an infrequently presented stimulus with a special meaning. Furthermore, the CIT is constructed in a way that people should pay attention to the stimuli of the task by requiring a special response to target items.

The P300 amplitude is the most prominent ERP analyzed in CIT studies (Leue & Beauducel, 2019; Meijer et al., 2014; Rosenfeld et al., 2013). Yet, some studies also analyzed MFN amplitudes for participants completing the CIT (Gibbons et al., 2018; Leue et al., 2012). They found enlarged MFN amplitudes for probe compared to irrelevant items (Gibbons et al., 2018; Leue et al., 2012). This pattern of MFN amplitudes revealed that denying knowledge of the probe items was accompanied by response conflicts. The concealment of knowledge of probe items seems to rely on additional cognitive control processes (Gibbons et al., 2018; Leue et al., 2012).

2.5 Moderators of ERPs in deception tasks

In many previous ERP-based deception studies, participants first commit a mock-crime and subsequently complete a CIT (Leue & Beauducel, 2019). Typically, participants are instructed to steal an object, e.g., a piece of jewelry or files (Gamer & Berti, 2012; Rosenfeld et al., 2018; Sai et al., 2020; Winograd & Rosenfeld, 2011). It is well documented that enlarged P300 amplitudes occur for crime-relevant probe items, compared to similar unknown irrelevant items (Leue & Beauducel, 2019; Rosenfeld et al., 2013). This line of work has motivated researchers to call for the future application of P300-based CITs in court (Farwell, 2012; Meixner, 2018; Rosenfeld et al., 2013). However, studies analyzing deception in contexts different from mock-thefts are scarce. A recent meta-analysis by Leue and Beauducel (2019) focused on the P300 in deception tasks. About 74% of the included studies with a pre-task scenario comprised an instructed mock-theft. It therefore remains questionable whether ERP patterns found in deception studies involving mock-thefts are generalizable to other situations, for instance more prosocial situations. Only 2

out of all 77 included studies in the meta-analysis focused on deception that could be seen as morally more acceptable or other-oriented: Lefebvre et al. (2007) investigated deception of witnesses rather than perpetrators and Hu et al. (2011) analyzed lies of self- and other-oriented information. There seems to be a research gap for deception about socially more acceptable behavior that needs to be bridged.

Likewise, Leue and Beauducel (2019) highlighted the importance of studying moderating effects of individual differences and sex on ERPs during deception. They were not able to analyze moderating effects of individual differences in their meta-analysis because there were not enough deception studies investigating them. Yet, in a research study, Leue et al. (2012) found a moderating effect of sensitivity to injustice on ERP components during a CIT. The difference of MFN amplitudes between probe and irrelevant items was larger for participants with higher sensitivity to injustice. Accordingly, deceptive responses for probe items were accompanied by stronger response conflicts for participants higher in sensitivity to injustice. Likewise, the difference of P300 amplitudes between probe and irrelevant items was moderated by sensitivity to injustice. Probe items were more salient for participants with higher sensitivity to injustice. In a following study, Leue and Beauducel (2015) analyzed combined effects of sensitivity to injustice and gender. Results were similar to their previous study. For women, sensitivity to injustice moderated the difference of P300 amplitudes between probe and irrelevant items. Probe items were more salient for women higher in sensitivity to injustice. However, the moderating effect of sensitivity to injustice did not occur for men. The findings were explained by a differing sensitivity to moral values depending on gender (Leue & Beauducel, 2015). For women, lying was probably seen as unjust and breaking of a social rule (Leue & Beauducel, 2015). In contrast, men possibly perceived lying as following the instruction of a task and did not consider the morality of lying (Leue & Beauducel, 2015). In conclusion, it may be important to consider the moral context when studying cognitive processes during deception. It is possible that differences in the moral context of deception coincide with a different cognitive processing of deception. The present thesis therefore focused on the moral context of the deception task, which can be altered by the situation but also interindividual differences. In the following, an introduction into moral theories will be given as well as aspects people consider when judging a behavior as moral or immoral.

2.5.1 The moral context

Moral theories address the question of what is right/good and what is wrong/bad (Ellemers et al., 2019; Mackey & Elvey, 2021). According to Turiel (1983), "The moral domain refers to prescriptive judgements of justice, rights, and welfare pertaining to how people ought to relate to each other." Turiel's definition already names the concrete content of morality, that is, justice, rights, and welfare. Judgements about them can be formed by experiences in social relationships when perceiving harm done to other people, violations of rights, and discussions about conflicting demands (Turiel, 1983). Due to moral virtues and principles, people can be inclined to display behavior that has no positive consequences for themselves (Ellemers et al., 2019). Instead, they possibly foresee positive consequences for others, the society, or communities (Ellemers et al., 2019; Haidt & Kesebir, 2008). The need of moral behavior for social life is at the heart of a definition by Haidt and Kesebir (2008). Instead of naming the content of morality as Turiel, they focus on its function, by writing that "Moral systems are interlocking sets of values, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible." The regulation of selfishness and social life can be obtained in multiple ways (Haidt & Hersh, 2001; Haidt & Kesebir, 2008). The exact content of these values, practices etc. is intentionally left open. Haidt and Kesebir (2008) wanted to take into account that their content can vary across cultures.

Two major moral theories are deontology and utilitarianism. Deontologists argue that acts in themselves are right or wrong, independent of their consequences (Haidt & Kesebir, 2008). The nature of the action defines whether people are ought to do it (Conway & Gawronski, 2013). An act that harms others is considered wrong, regardless of whether it helps more people in total (Conway & Gawronski, 2013). Immanuel Kant created the most influential deontological theory (Haidt & Kesebir, 2008; Mackey & Elvey, 2021). In his categorical imperative, he outlined that people should act after rules that are generalizable to universal rules for everyone (Kant, 1785). Actions are judged as right when a person consistently and rationally would want everyone to follow a rule determining the action (Haidt & Kesebir, 2008; Kant, 1785).

Conversely, for utilitarians not the inherent righteousness of an action is decisive but its consequences. Actions are considered morally right when they bring the most good

for the most people (Driver, 2014; Mackey & Elvey, 2021). Jeremy Bentham and John Stuart Mill were two key advocates of utilitarianism, setting its ground foundations (Driver, 2014; Mackey & Elvey, 2021). They considered good what brings pleasure and reduces pain (Bentham, 1789/1961; Driver, 2014; Mackey & Elvey, 2021; Mill, 1863/2001). Accordingly, one is ought to do what promotes the greatest happiness for the greatest number of people (Bentham, 1789/1961; Mill, 1863/2001).

Despite their differences, many moral theories outline that protecting, helping or benefitting others is essential for morality (Gilligan & Attanucci, 1988; Gray & Wegner, 2011; Haidt & Graham, 2007; Kohlberg, 2008). People should consider the welfare/happiness of others and avoid harming them (Gilligan & Attanucci, 1988; Gray et al., 2012; Gray & Wegner, 2011; Haidt & Kesebir, 2008; Kohlberg, 2008). Likewise, people judge the reduction of harm as central to morality (Graham et al., 2011). In a study by Graham et al. (2011), 34,476 participants gave ratings about the moral relevance of harm doing and other aspects possibly relevant for morality. Graham et al. (2011, p. 15) came to the conclusion that "Concerns about harm and fairness are so widespread that they might be said to be universally used foundations of morality (upon which cultures construct differing ideas as to what counts as harm or what kinds of distributions are fair)." Relatedly, in the literature, moral behavior is often equated with helping, altruism, and prosocial behavior (Doris et al., 2018; Haidt & Kesebir, 2008). For instance, Rachels (2013) states, "Moral behavior is, at the most general level, altruistic behavior, motivated by the desire to promote not only our own welfare but the welfare of others." The term *prosocial behavior* subsumes actions that are intended to benefit others (Batson & Powell, 2003; Eisenberg et al., 2006). The motivation for prosocial behaviors can be manifold. Conversely, altruistic behavior refers to behaviors with the primary motivation, or in other words the ultimate goal, to increase the welfare of others (Batson, 1987; Batson & Powell, 2003). In accordance with the focus of moral theories on promoting vs. impairing the welfare of others, many studies about morality investigate actions that are intended to harm vs. help people (Decety et al., 2015; Kenward & Dahl, 2011; Leslie et al., 2006). This line is also followed in the present thesis. Nevertheless, it should be noted that some theories outline additional aspects of morality (see Haidt, 2008; Haidt & Kesebir, 2008; Killen & Smetana, 2014).

2.5.1.1 Deception in different moral contexts

"Moral decency ensures for us the right to be deceived as surely as the right to truth: to extol the latter and deny the former is to misunderstand being human."

— David Nyberg, The varnished truth, 1993

According to Kant, it is everyone's duty to always speak the truth, even in situations in which a lie possibly prevents harm (Kant, 1797). He outlines that lies result in an untrustworthiness of statements in general, and therefore cause injustice to mankind (Kant, 1797). In contrast, people lie frequently in their everyday life (Ennis et al., 2008; Kashy & DePaulo, 1996). People tell lies for various reasons, which can be self- but also other-oriented (Camden et al., 1984; DePaulo & Bell, 1996; Turner et al., 1975). In a diary study by DePaulo et al. (1996), about 25% of the lies were prosocial (meaning lies told with the intent to benefit others). Furthermore, people also lie about prosocial behavior in order to protect others' self-worth or avoid conflicts (Camden et al., 1984; DePaulo & Kashy, 1998; Turner et al., 1975). There is even a term for harmless lies: White lies are generally considered socially accepted lies (Bryant, 2008; Camden et al., 1984; D'Agata, 2014). They are often called trivial in that they have little consequences and rather the potential to avoid harm (Bryant, 2008; Camden et al., 1984; D'Agata, 2014; Sweetser, 1987). White lies are common in social interactions (Camden et al., 1984; DePaulo & Bell, 1996; DePaulo & Kashy, 1998; Turner et al., 1975). In contrast to Kant's perspective on lying, for utilitarians the consequences expected to follow a lie are decisive for its moral value (T. L. Carson, 2010). From the utilitarian perspective, a lie can be judged as morally permissible when it leads in sum to more happiness or reduction of pain than any other act (T. L. Carson, 2010). However, in order to be consistent, one needs to assume that the untrustworthiness of statements in general that might even follow from white lies, does not result in a substantial increase of harm for a large number of individuals.

After this short look at lying from the perspective of different philosophical theories, the question remains how people perceive lying. In general, people have a negative view on lying (Backbier et al., 1997; Lavoie et al., 2016; Robinson, 1994). Yet, when judging different kind of lies, they rate some lies as more acceptable than others (Levine & Schweitzer, 2014; Lindskold & Han, 1986; Robinson, 1994). Levine and Schweitzer (2015) found that people see prosocial lies as morally permissible. In their study, people

even judged prosocial lies as morally superior to selfish truths. Likewise, other studies found that people rate prosocial lies as more acceptable than selfish lies and indicate feeling less remorse when telling them (Hayashi et al., 2014; Lavoie et al., 2016; Lindskold & Han, 1986; McLeod & Genereux, 2008; Peterson, 1996; Robinson, 1994; Seiter et al., 2009; Seiter & Bruschke, 2007).

Independent of the objective definition of a lie (see Chapter 2.1), the subjective perception can differ between prosocial and egoistic lies. When judging to which extent a statement is perceived as a lie, people take its harm vs. helpfulness into account (Lee & Ross, 1997; Peterson et al., 1983; Xu et al., 2009). This is in line with Sweetser's folkloristic model (1987) saying that the context has to be considered when defining lies. Giving somebody truthful information is normally beneficial for the recipient, in contrast, lies are usually harmful (Sweetser, 1987). However, this is not the case for prosocial lies (Sweetser, 1987). In accordance with Sweetser's model, studies by different research groups found that people judge lies intended to promote the welfare of others as less representative of lies than those told out of selfish reasons (Lee & Ross, 1997; Peterson et al., 1983; Xu et al., 2009). These different perspectives on prosocial and other types of lies also illustrate that it seems worthwhile to study whether cognitive processes differ between lies with a more prosocial or egoistic context. It should yet be noted that the extent to which people rate prosocial lies as representative for lies is out of the scope of the thesis.

2.5.1.2 ERP studies with different moral contexts

ERP studies analyzing deception in different moral contexts are scarce. However, some studies found moderating effects of variables related to the moral context on ERP amplitudes. In the meta-analysis by Leue and Beauducel (2019), higher effect sizes were found for the difference of P300 amplitudes between deceptive and truthful answers when deception was studied in a legal compared to a social context. Studies with legal context mainly comprised mock crimes. Conversely, studies with social context comprised deception about attitudes, concealing knowledge about autobiographical data, or card games. Accordingly, it could be argued that these categories of studies also differed in their moral context. Studies with a legal context involved situations that could be perceived as immoral, whereas the context of studies with social situations could rather be

perceived as neutral.

Some studies compared the P300-based CIT effect for culprits of a mock-crime with witnesses or participants who were informed about the mock-crime (Jang et al., 2013; Rosenfeld, Ozsan, et al., 2017; Winograd & Rosenfeld, 2014). Culprits performed or imagined performing an immoral behavior, like stealing a ring or intentionally hitting another person in a car crash. Conversely, witnesses and informed innocents did not perform an immoral behavior. They saw somebody else performing the immoral behavior or were informed about it. Hence, the two groups differed in their involvement in the behavior causing harm to other people. Therefore, here and in the following, I refer to *effects of moral involvement* when comparing people who are actively involved in a behavior causing harm/problem and people who are not actively involved in the behavior (witnesses/informed people). Moral involvement should not be mixed up with moral disengagement, a term introduced by Bandura (1986). Moral disengagement does not refer to the involvement of a person in a certain behavior but subsumes cognitive mechanisms used to disengage from self-sanctions of immoral behavior, e.g., moral justification or diffusion of responsibility (Bandura et al., 1996; Moore et al., 2012). In existing studies, P300-based CIT effects were found for both culprits and witnesses/informed participants (Jang et al., 2013; Lefebvre et al., 2007, 2009; Rosenfeld, Ozsan, et al., 2017; Winograd & Rosenfeld, 2014). However, results regarding the moderating effect of moral involvement were mixed. Winograd and Rosenfeld (2014) investigated P300s during a CIT for culprits and informed innocents. Culprits got instructions for a mock-theft and subsequently committed it. Informed innocents were informed about details of the mock-theft and the identity of the probe item by the same instructions but did not commit the theft. The P300-based CIT effect did not differ between the two groups. Jang et al. (2013) compared P300s for three groups of participants: Culprits committing a crime in a virtual reality, witnesses observing the crime in a virtual reality, and informed participants, who read about the crime and were shown pictures of it. P300 amplitudes in a following CIT did not differ between culprits and witnesses. Yet, P300 amplitudes of probe items were reduced for informed participants, compared to culprits and witnesses. Lastly, in a study by Rosenfeld, Ozsan, et al. (2017), participants read a newspaper article and a video of a theft. They were instructed to imagine that they either committed or witnessed the theft. Larger differences of P300 amplitudes between probe and irrelevant items occurred for

culprits than for witnesses. The authors concluded that probe items were more meaningful for culprits, and they therefore turned their attention more towards probe items.

Similarly, Lu et al. (2018) analyzed P300 amplitudes during a CIT for participants who committed a mock-theft alone and participants who committed it together with another person. P300 amplitudes were enlarged for probe compared to irrelevant items for both groups. Nevertheless, the difference of P300 amplitudes between probe and irrelevant items were larger for participants who committed the mock-theft alone, compared to participants who committed it together with another person. Moreover, it has also been shown that greater differences of P300 amplitudes between probe and irrelevant items occur for people with a higher awareness of deception (Rosenfeld et al., 2012). Table 1 gives an overview of the CIT studies analyzing the effect of moral variables on the P300. Even though the study results regarding moderating effects of variables related to the moral context on the P300-based CIT effect are mixed, findings of some research studies suggest that probe items could be perceived as more meaningful in immoral contexts leading to an enlarged P300-based CIT effect. Therefore, it might be worthwhile to further investigate the moderating effect of variables related to morality on the P300-based CIT effect. This might also help to further understand and find explanations for the differing results of previous studies.

Regarding moderating effects on MFN amplitudes, recall the study by Leue and Beauducel (2019). They found an increased difference of MFN amplitudes between deceptive responses for probe items and honest responses for irrelevant items for people with a higher sensitivity to injustice. There is, to the best of my knowledge, no other study analyzing the moderating effect of variables related to morality on MFN amplitudes during deception. However, some research groups analyzed MFN amplitudes during moral judgements. In a study by Boksem and De Cremer (2010), MFN amplitudes were analyzed during an ultimatum game. During the game, participants received unfair monetary offers, in which the sender of an offer assigned himself much more money than the participant, as well as fair monetary offers, in which the sender of an offer assigned himself and the participant the same amount of money. MFN amplitudes were enlarged for unfair compared to fair offers. Moreover, a larger difference of MFN amplitudes between fair and unfair offers occurred for participants with higher concerns of fairness. Likewise, enlarged MFN amplitudes were found for fair compared to unfair offers in a study by Hu

and Mai (2021). They also reported a larger difference of MFN amplitudes between fair and unfair offers for individuals with a prosocial orientation (i.e., individuals focusing on the maximization of joint outcomes even at the expense of their own outcomes), compared to individuals with a pro-self orientation (i.e., individuals focusing on the maximization of their own outcomes even at the expense of the outcomes of others). An important neural source of the MFN, the ACC, has also been associated with moral conflicts in some studies (Greene et al., 2004; Parkinson et al., 2011). Increased activity of the ACC has been found for people faced with difficult moral dilemmas (Greene et al., 2004). Similarly, an increased activity of the ACC has been reported for people given descriptions of moral transgressions involving physical harm, compared to descriptions of neutral scenarios (Parkinson et al., 2011). Altogether, the result of these studies suggest that the ACC and the MFN may also be sensitive to moral conflict.

Table 1

Deception studies analyzing effects on variables related to morality on P300 amplitudes

Study	Groups	Manipulation/Task	CIT effect	Moderating effect
Jang et al. (2013)	Culprits (*n* = 15) Witnesses (*n* = 15) Informed innocents (*n* = 15)	In a VR, culprits crashed their car into another car; witnesses observed the car crash; informed innocents read about it	Yes	No differences in P300s between witnesses and culprits. Larger P300s of probes for culprits than informed innocents.
Rosenfeld, Ozsan, et al. (2017)	Culprits (*n* = 15) Witnesses (*n* = 16)	Participants read an article and saw a video of a theft. They imagined witnessing or committing the theft.	Yes	CIT effect occurred in both groups. The CIT effect was larger for culprits than witnesses.
Winograd & Rosenfeld (2014)	Informed innocent (*n* = 13) Informed guilty (*n* = 16)	In both groups, participants got instructions for a theft. Only guilty participants committed it.	Yes	No differences in P300s between the two groups
Lu et al. (2018)	Individual group (*n* = 18) Collaborative group (*n* = 18)	Participants committed theft alone or with another person.	Yes	CIT effect occurred in both groups. The CIT effect was larger in the individual than the collaborative group.
Rosenfeld et al., (2012)	Control group (*n* = 12) Deception group (*n* = 12)	The CIT involved towns, inter alia, the hometown of the participant. The control group was instructed to sort the towns as targets and non-targets. The deception group was instructed to indicate whether they know the towns and respond deceptively for their hometown.	Yes	Larger CIT effect in the deception than the control group

Note. VR= virtual reality; CIT effect= Difference of P300s between probe and irrelevant items. The study by Winograd and Rosenfeld (2014) involved additional groups that are not named in the table, since they are not relevant for comparisons of the moral context.

2.5.2 Machiavellianism

"[F]or a long time I have not said what I believed, nor do I ever believe what I say, and if indeed sometimes I do happen to tell the truth, I hide it behind so many lies that it is hard to find."

— Niccolò Machiavelli, Letter to Francesco Vettori, 1521

The importance of considering personality traits when studying deception has been repeatedly emphasized by Leue and colleagues (Leue et al., 2012; Leue & Beauducel, 2015, 2019). A personality trait that has been closely related to deception and to a disregard of conventional morality is Machiavellianism (Christie & Geis, 1970; Fehr et al., 1992; Jones & Paulhus, 2009). Christie and Geis (1970) conceptualized Machiavellianism based on the ideas of Machiavelli in his book *the Prince.* Machiavelli was a political advisor in the 16th century (Jones & Paulhus, 2009). In *the Prince,* he outlined that rulers should not shy back from immoral behavior when needed to maintain power, true to the motto "the end justifies the mean" (Jones & Paulhus, 2009; Machiavelli, 2009). Christie and Geis (1970) proposed that people differing in their agreement to principles laid out by Machiavelli also display stable differences in their behaviors in everyday life. They introduced the personality trait Machiavellianism.

Machiavellianism is defined by (1) the use of manipulative and strategical tactics, such as deception in interpersonal relationships (Christie & Geis, 1970; Fehr et al., 1992). A callous affect and lack of empathy helps Machiavellians to view others as manipulable and to use manipulative tactics for persuading other people to help them reach their goals (Christie & Geis, 1970). Individuals high in Machiavellianism are more successful in staying detached from irrelevant emotional information and instead focus on their goals (Geis et al., 1970). In a study by Geis et al. (1970), individuals with low vs. high Machiavellianism scores did not significantly differ in succeeding at a game comprising neutral issues. However, when the same game involved emotional issues individuals low in Machiavellianism seemed to get distracted from the emotional content and were less successful than individuals high in Machiavellianism (Geis & Moon, 1981). A further characteristic of Machiavellianism is (2) a cynical worldview. Other people are seen as selfish, weak, and untrustworthy (Christie & Geis, 1970; Fehr et al., 1992). Finally,

Machiavellianism is characterized by (3) a lack of conventional morality. Individuals high in Machiavellianism have a pragmatic view on relationships and on behaviors others normally view morally reprehensible, such as lying and cheating (Christie & Geis, 1970). They more likely engage in unethical behavior (Hegarty & Sims, 1978, 1979), rate ethical questionable behavior as more acceptable (Mudrack, 1993; Mudrack & Mason, 1995), and indicate to experience fewer moral conflicts in face of unethical behavior (Mudrack & Mason, 1995). The conceptualization of Machiavellianism was also influenced by the Chinese military strategist Sun Tzu (Jones & Paulhus, 2014). Based on his writings, Jones and Paulhus (2014) added planning, coalition formation, and reputation building as characteristic of Machiavellianism. They highlighted that Machiavellians plan ahead strategically and take care of their reputation (Jones & Paulhus, 2014). In accordance, individuals scoring high on Machiavellianism help others more frequently when they are observed by group mates and they can hereby increase their reputation in a group (Bereczkei et al., 2010).

2.5.2.1 Machiavellianism and deception

Machiavelli saw lying as a permissible mean to reach one's goals (Machiavelli, 2009). In this line, multiple research groups found correlations between Machiavellianism and deception: Individuals high in Machiavellianism indicate to lie more frequently in everyday life (Azizli et al., 2016; Kashy & DePaulo, 1996). Likewise, they are more prone to lie for financial gain in a laboratory setting (Ghosh & Crain, 1995; Murphy, 2012; Sakalaki et al., 2007), and have lower intentions to stay true to arranged deals (Forgas, 1998). Furthermore, Machiavellianism correlates negatively with the factor honesty/humility of the hexaco model as well as with its facet sincerity (*r* about -.40; Ashton et al., 2000; Lee & Ashton, 2005). Likewise, associations have been found between Machiavellianism and the telling of multiple types of lies. Individuals high in Machiavellianism indicate to tell more white lies with the intention not to hurt another person than individuals with lower Machiavellianism scores (Jonason et al., 2014). They also more likely tell self-serving lies as well as lies intended to avoid conflicts (Jonason et al., 2014; McLeod & Genereux, 2008). Azizli et al. (2016) found that all personality traits of the dark triad (i.e., Machiavellianism, narcissism, and psychopathy) correlate with the propensity for high-stakes deception. Yet, in a multiple regression considering the dark triad traits and the

participant's gender, only Machiavellianism was a significant predictor of high-stakes deception (Azizli et al., 2016). Machiavellians also seem to have fewer difficulties with telling lies: When thinking about the production of lies, individuals higher in Machiavellianism perceive lying as cognitively less challenging (Gozna et al., 2001), and rate their abilities in lying as higher (Gozna et al., 2001; Wissing & Reinhard, 2019). They also indicate to feel less guilty when lying, compared to individuals lower in Machiavellianism (Gozna et al., 2001; Murphy, 2012). In a nutshell, individuals high in Machiavellianism seem to have less scruple to lie.

2.5.2.2 ERP studies about Machiavellianism and deception

Even though Machiavellianism has been closely related to deception, little is known about the moderating effect of Machiavellianism on P300 or MFN amplitudes during deception. Panasiti et al. (2014) found an association between Machiavellianism and the Bereitschaftspotential (i.e., a slow negative potential preceding a response) during deception. In their study, participants were honest as well as deceptive about the outcome of a card game. The amplitude of the Bereitschaftspotential was reduced for deceptive compared to honest answers. For participants scoring lower on Machiavellianism, the Bereitschaftspotential of deceptive responses was more reduced. It has to be specified that responses were the same for deceptive and honest responses. Modulations of the Bereitschaftspotential could therefore not be explained by different response types. Panasiti et al. (2014) connected the modulations of the Bereitschaftspotential to moral conflicts and moral dilemmas. Accordingly, moral conflicts were increased for deceptive compared to honest responses. Likewise, moral conflicts during deception were stronger for people lower in Machiavellianism. Furthermore, in an fMRI study, Fullam et al. (2009) found associations between Machiavellian Egocentricity (which is characterized by looking out for one's own interests before others) and the brain circuit that is activated during deception.

A trait that also belongs to the dark triad and is related to Machiavellianism is psychopathy. Research findings regarding the P300 and psychopathy are mixed. As already outlined, the CIT can be seen as a kind of oddball task (see Chapter 2.4). In some studies using oddball tasks, reduced P300 amplitudes were found for psychopathic compared to nonpsychopathic individuals (Kiehl et al., 1999, 2006). In contrast, in other

studies, P300 amplitudes of oddballs were enlarged for psychopaths (Raine & Venables, 1988), and there are also studies in which no difference in P300 amplitudes between psychopathic and nonpsychopathic individuals occurred (Campanella et al., 2004; Jutai et al., 1987). However, in a meta-analysis, an overall reduced P300 amplitude was found for psychopaths in standard oddball tasks (Gao & Raine, 2009). A. R. Miller and Rosenfeld (2004) analyzed P300 amplitudes during a CIT for individuals categorized as highly and less psychopathic. P300 amplitudes did not significantly differ between the two groups. However, the authors discussed that effects of psychopathy were possibly not seen in their study due to opposing gender effects. The group of highly psychopathic individuals consisted mostly of men, whereas the group of less psychopathic individuals consisted mostly of women.

2.6 Goals of the thesis

Existing deception studies have a strong focus on finding the best way to detect deception (Rosenfeld, 2020; Rosenfeld et al., 2013). Deception tests and methods for the processing of EEG data are optimized for applied forensic assessments. In this regard, the CIT has prevailed as a method for detecting deception in the form of concealed knowledge, especially for mock-thefts (Rosenfeld et al., 2013). Concealed knowledge about objects from the theft is typically revealed by patterns of P300 amplitudes. Previous CIT studies already compared methods for the detection of concealed knowledge of mock-thefts, for instance, the quantification of ERP amplitudes and the way of the presentation of the stimuli of the CIT, in order to increase the number of accurately detected people who concealed knowledge (Ambach et al., 2010; Rosenfeld et al., 2015; Soskins et al., 2001). However, to ensure a successful application of deception tests, they also have to be based on a profound theory. It seems mandatory that possible determinants of the ERP patterns during deception are known. Confounding variables have to be revealed. This approach opens the possibility to learn more about the cognitive processes involved in deception tasks. Therefore, the goal of the present thesis is to advance the theoretical basis of deception tasks by investigating determinants of ERP patterns during their completion. This line of research can help to set up a theoretical nomological net of the cognitive processes during deception tasks.

The P300 is an ERP that has commonly been analyzed in deception tasks.

Different patterns of P300 amplitudes for deceptive compared to honest answers have repeatedly been found, especially for studies involving mock-crimes (Leue & Beauducel, 2019). In contrast, the investigation of MFN amplitudes during deception is in its infancy and there are–to the best of my knowledge–only a handful of deception studies analyzing patterns of MFN amplitudes (Gibbons et al., 2018; Johnson et al., 2004, 2008; Leue et al., 2012; Suchotzki et al., 2015). Therefore, one aim of the current thesis was to investigate whether response conflicts, as indicated by MFN amplitudes, commonly occur during deception. As already outlined, deception studies often involve forensic scenarios. In the studies of the current thesis, I wanted to focus on situations apart from classic mock-thefts. It remained questionable whether the difference of P300 and MFN amplitudes between deceptive and honest responses also relies on the feature that they are usually analyzed in forensic scenarios. In order to get to know whether the salience of the probe item and the conflicts during deception are not based on the fact that the probe items are typically seen during a crime scene, these ERP components also had to be analyzed in situations apart from mock-crimes. Furthermore, situations people lie about are manifold. Investigating P300 and MFN amplitudes also in non-forensic settings can further the understanding of the cognitive processes during deception. Therefore, all studies of the thesis involved at least one condition with a social, non-forensic situation. I was interested whether different patterns of P300 but also MFN amplitudes occur repeatedly in studies involving other situations than mock-thefts. Thereby, the generalizability of the patterns of P300 and MFN amplitudes can be tested. Likewise, this allows to unravel the ERP's determinants. Accordingly, the first research question, investigated in all three studies of the thesis, was the following.

1. *Research question: Do different patterns of P300 as well as MFN amplitudes occur for deceptive compared to honest responses?*

Since enlarged MFN amplitudes indicate conflicts and deceptive responses should be accompanied by more intense response conflicts than being honest, I expected enlarged MFN amplitudes for deceptive compared to honest responses. The P300 has a dual nature in different deception paradigms (see Chapter 2.3.5): CIT studies rely on the feature of the P300 to indicate salience of infrequently presented, meaningful stimuli in a series of frequently presented, irrelevant stimuli. For a person who recognizes the meaningful

(probe) stimulus, it should appear more salient than other unknown irrelevant stimuli even when the recognition is denied. Therefore, the possibility to reveal concealed knowledge by the pattern of P300 amplitudes emerges. In the CIT studies of the present thesis (study I and III), participants gave deceptive responses for probe items, in that they indicated that they do not know them, although they had seen them before. For irrelevant items, they indicated honestly that they do not know them. Since probe items should appear more salient, I expected enlarged P300 amplitudes for probe compared to irrelevant stimuli. In studies involving deception tasks with an equal frequency of honest and deceptive responses and which do not rely on the recognition of items, suppressed P300 amplitudes were found for deceptive compared to honest responses (Dong et al., 2010; Johnson et al., 2008; Meek et al., 2013; Pfister et al., 2014). Since items requiring a deceptive response are frequently presented and do not represent known items in a series of unknown irrelevant items, they should not appear more salient than items requiring honest responses. Instead, responding deceptively should be cognitively more challenging, should capture cognitive resources, and therefore drive attention away from stimulus processing. Correspondingly, for deception tasks apart from the CIT, suppressed P300 amplitudes were expected for items requiring a deceptive response, compared to items requiring an honest response (study II).

Furthermore, in the studies of the thesis, I aimed to test whether cognitive processes during deception tasks are similar for different individuals and situations, or if certain moderator variables have to be considered. Some previous studies found that variables related to the moral context may moderate ERPs during deception (Leue et al., 2012; Leue & Beauducel, 2015; Rosenfeld et al., 2012; Rosenfeld, Ozsan, et al., 2017). Therefore, a special focus was set on variables related to the evaluation of the moral context of the deception task. The evaluation of the moral context can be determined by individual differences but also situational variables. Regarding individual differences, a trait that seems of special interest is Machiavellianism. Since individuals high in Machiavellianism seem to have less scruple to lie (see Chapter 2.5.2.1), it is possible that they also process lying differently than individuals low in Machiavellianism. Therefore, an additional research question of the thesis (investigated in studies I and II) was:

2. *Research question: Is the difference of MFN and P300 amplitudes between deceptive and honest responses moderated by Machiavellianism?*

Individuals high in Machiavellianism are more prone to lie (Azizli et al., 2016; Kashy & DePaulo, 1996). Furthermore, they indicate to feel less guilty when lying and that lying is cognitively less strenuous for them (Gozna et al., 2001; Murphy, 2012). I wanted to explore whether these reports in questionnaire and diary studies are also reflected in the cognitive processes indicated by ERP amplitudes during deception tasks. Accordingly, I expected that deception would be accompanied by less intense conflicts for individuals higher in Machiavellianism, as indicated by a smaller difference of MFN amplitudes between deceptive and honest responses. Furthermore, I expected that deception would be cognitively less challenging for them. In deception paradigms beyond the CIT, in which participants have to lie equally often as respond honestly, suppressed P300 amplitudes for lies, compared to honest responses, reflect the additional mental workload for lying. Since lying should be cognitively less challenging for individuals higher in Machiavellianism, I expected a smaller difference of P300 amplitudes between deceptive and honest responses for them in these deception paradigms.

Situations people lie about and behaviors they try to cover up can be morally less or more reprehensible. For instance, people can cover up behavior of other people to save them from criticism or lie about a prosocial behavior, which could be judged as morally less reprehensible than lying about theft. It has already been found that MFN amplitudes can be sensitive to moral conflict (Boksem & De Cremer, 2010; Leue et al., 2012). When a lie is told in a situation in which the wellbeing or happiness of another person is in focus, lying can be even considered the morally superior way than being honest (Levine & Schweitzer, 2015). Accordingly, people possibly perceive less (moral) conflict when lying in a prosocial context, leading to different patterns of MFN amplitudes. Furthermore, Leue and colleagues found that moral sensitivity, represented in sensitivity to injustice, moderates the patterns of P300 amplitudes during the CIT (Leue et al., 2012; Leue & Beauducel, 2015). The question arose whether such a moderation effect can also be found for variables beyond personality traits, such as the morality of the situation in which the probe item is seen. The third research question, investigated in studies I and III, was the following.

3. *Research question: Is the difference of MFN and P300 amplitudes between deceptive and honest responses moderated by situational variables related to morality?*

I expected smaller differences of MFN amplitudes between deceptive and honest responses in more prosocial situations. Lying was expected to be accompanied by less (moral) conflict in such situations. Furthermore, I aimed to investigate whether probe items requiring a deceptive response are less significant when they are seen in a prosocial context. This could lead to a smaller difference of P300 amplitudes between probe and irrelevant items.

Altogether, the investigation of moderating effects on MFN and P300 amplitudes during deception tasks helps to build a conceptual framework of the cognitive processes during deception. If some variables moderate the ERP patterns during deception, these variables could explain inconsistencies in seemingly contradicting findings of deception studies. For instance, even though enlarged MFN amplitudes usually occur for deceptive compared to honest responses (Johnson et al., 2005, 2008; Leue et al., 2012), Suchotzki et al. (2015) found enlarged MFN amplitudes for honest compared to deceptive answers. Suchotzki et al. (2015) discussed that participants in their study possibly perceived lying as a positive and therefore correct response, since lying was promoted in their study. This would mean that the contextual embedding of lies and how they are perceived by participants are also decisive for the patterns of MFN amplitudes. When such moderators of ERP components during deception paradigms are revealed, they can be considered in future studies, producing results that might be more comparable. Overall, this would lead to a clearer picture of the processes at work during deception.

To investigate the research questions, three studies were conducted, which are listed in Table 2. In study I, individual differences in ERP patterns during CIT were investigated. We analyzed the moderating effect of personality characteristics, with a special focus on Machiavellianism. Moreover, the moderating effect of moral involvement was investigated. In an uninstructed scenario before the CIT, participants either performed a behavior causing a social conflict or witnessed the same behavior. In study II, ERPs were investigated for a deception paradigm that did not involve the concealment of knowledge. We aimed to replicate the results of a study by Johnson et al. (2008), in which participants lied about their attitudes. Attitude evaluations also touch moral themes in a wider sense as they are derived from a person's values. As Johnson et al. (2008, p. 470) put it, "A central aspect of attitude evaluations is the requirement to make affective judgement about the goodness/badness of the attitude object based on internal scales reflecting

the individual's personal value system." It is expected that the individual personal values are closely related to moral social norms (cf. Rokeach, 1973; Weber, 1993). Moreover, again, moderating effects of Machiavellianism on P300s and MFNs in this deception paradigm were analyzed. Finally, in study III, the moderating effect of the moral valence on the P300- and MFN-based CIT effect was investigated: Participants concealed knowledge of an item seen during a pro- vs. antisocial behavior. As a side note: Besides the fact that all studies investigate ERPs during deception tasks with a focus on different moral contexts, a further binding element of them is that variables related to the involvement of the participant play a central role. In study I, moderating effects of moral involvement are investigated, meaning in this case witnessing or demonstrating a behavior causing a small social problem. Furthermore, a characteristic of Machiavellianism is a form of non-involvement in moral topics. Lastly, attitudes and the valence of a stimulus can trigger the involvement of a person in a situation or behavior. The overarching perspective of the three studies, investigating moral involvement (witnessing vs. performing behavior), Machiavellianism, and moral valence (pro- vs. antisocial behavior), is that this may allow for the detection of processes beyond the recognition of known probes. The investigated variables may all modulate the intensity of deception processes and may therefore affect the P300 and MFN amplitudes.

Table 2

Studies of the thesis

Study	Reference
I	**Scheuble, V**., & Beauducel, A. (2020). Individual differences in ERPs during deception: Observing vs. demonstrating behavior leading to a small social conflict. *Biological Psychology*, *150*, 107830. https://doi.org/10.1016/j.biopsycho.2019.107830
II	**Scheuble, V.**, & Beauducel, A. (2020). Cognitive processes during deception about attitudes revisited: a replication study. *Social Cognitive and Affective Neuroscience*, *15*(8), 839-848. https://doi.org/10.1093/scan/nsaa107
III	**Scheuble, V.**, Mildenberger, M. & Beauducel, A. (2021). The P300 and MFN as indicators of concealed knowledge in situations with negative and positive moral valence. *Biological Psychology*, *162*, 108093. https://doi.org/10.1016/j.biopsycho.2021.108093

3 Method

In the following, the general methods of the studies will be described, meaning the recording, processing and analyses of EEG data. Specific details of the methods of the research studies can be found in the corresponding articles, and a general description of the applied deception tasks are also given in the chapters of the research studies. Two studies in the thesis (studies I and III) investigated deception in a CIT. For CIT studies, certain data preparation and quantification methods of the P300 have prevailed, especially since they have proven effective for detecting concealed knowledge. Pre-processing steps of EEG data of these studies were chosen in accordance to previous CIT studies to ensure the comparability of their results (Leue et al., 2012; Rosenfeld, Ozsan, et al., 2017). Conversely, study II did not analyze deception in a CIT but in a task in which participants lied about their attitudes. Study II was a replication study. The methods of this study were chosen in line with the original study by Johnson et al. (2008) to minimize the possibility that differences between the results of the replication and original study are attributable to differing data preparations. Table 3 at the end of Chapter 3.3 gives an overview of the settings of the data preparation steps in each of the three studies. The following descriptions of the methods apply to all three research studies, unless otherwise stated.

3.1 EEG recording and pre-processing

The recorded potential at the scalp is not only determined by the brain activity of interest but also influenced by other electrical signals, stemming for example from muscular activity, blinks, and skin potential (Fabiani et al., 2017; Luck, 2014). Such artifacts have to be minimized during recording and pre-processing of EEG data to optimize the signal-to-noise ratio (Fabiani et al., 2017; Luck, 2014). To reduce the possibility that electrical noise of surrounding devices contaminates the EEG signal, the EEG examinations took place in an electrically shielded room. The EEG was recorded by 64 Ag/AgCl active scalp electrodes (Active Two BioSemi, Amsterdam, Netherlands), which were placed according to the extended 10-20 system (Jasper, 1958). The 10-20 system is a standardized electrode placement, ensuring that EEGs of different laboratories are comparable (Milnik, 2012). Additional electrodes were placed near the eyes to record eye movements and to correct and eliminate such artifacts contaminating the data (the specific placement of

EOG electrodes is given in Table 3). The EEG signal is very weak and therefore needs to be amplified so that its trajectory is recognizable to the naked eye (Luck, 2014). However, the amplifier generates additional electrical noise (Luck, 2014). Ground electrodes serve to eliminate such electrical noise from the EEG signal (Luck, 2014). In contrast to other EEG systems with one ground electrode, the Biosemi system has two ground electrodes, the common mode sense (CMS) active electrode and driven right leg (DRL) passive electrode. The CMS and DRL electrode form a feedback loop. The DRL electrode injects a small amount of current in the head, so that the average potential of the examinee is similar to the potential of the amplifier's ground circuit (Luck, 2014; *What Is the Function of the CMS and DRL Electrodes*, n.d.). This improves the rejection of electrical noise by the CMS electrode (Luck, 2014; *What Is the Function of the CMS and DRL Electrodes*, n.d.). The signal was sampled and digitized with ActiView (Biosemi). Electrode offsets are given in Biosemi systems to indicate contact problems between electrodes and the scalp as well as electrode corrosion (Smith, 2007). In line with the operating guidelines of ActiView (Smith, 2007, p. 51), electrode offsets were therefore kept below 30 mV.

Offline analyses, such as offline re-referencing, were performed with EEGLab (version 12.0.2.6b; Delorme & Makeig, 2004). As outlined in Chapter 2.3.1, voltage represents the potential of charged particles to move from one place to another (Luck, 2014). Therefore, a reference point, or in other words a reference electrode, is needed to measure the voltage at a current electrode. In study I and III, data were re-referenced to the averaged signal at electrodes P9/P10. They are like the references from previous CIT studies near the mastoids (Leue et al., 2012; Leue & Beauducel, 2015) but have the advantage that they do not record as much muscle activity from the neck as electrodes placed directly at the mastoids (Luck, 2014). In the replication study (study II), the same reference electrodes were chosen as in the original study by Johnson et al. (2008), that is, we used the averaged signal at pre-auricular sites. Accordingly, in all studies, average signals of electrodes placed at each hemisphere were used as references, avoiding a biased signal towards one hemisphere (Luck, 2014). Furthermore, we avoided that reference electrodes were near the place where the ERPs of interest are largest: at fronto-central and parietal-central electrodes.

Data were filtered offline, after the recording concluded. Filtering EEG data helps to reduce artifacts that can be approximated by sine waves (Luck, 2014). ERPs have

usually a frequency spectrum between approximately 0.1 and 30 Hz (Luck, 2014). Conversely, very low frequencies typically stem from noise by the electrodes or skin and very high frequencies typically stem from muscular activity rather than brain activity (Luck, 2014). Following previous CIT studies, the data from MFN amplitudes were band-pass filtered between 0.3 and 30 Hz in study I and III (Leue et al., 2012). Whereas MFN amplitudes are relatively small and narrow and a more restrictive low pass filter could distort MFN amplitudes, P300s have larger amplitudes and are more long-lasting. In accordance with filter settings of current P300-based CIT studies, a low pass filter of 6 Hz was applied on P300 data of study I and III (Olson et al., 2019; Rosenfeld, Ozsan, et al., 2017; Ward et al., 2020). This smoothens the curve of the P300, grand averages of different conditions can be more easily compared and, more importantly, results are better comparable to previous CIT studies. Following Johnson et al. (2008), the same band-pass filter was used for all ERPs in the replication study (study II), which was a bit less restrictive (0.1-35 Hz).

Ocular artifacts were corrected in study I and III by means of an independent component analysis (ICA with infomax decomposition; Leue et al., 2012; Leue & Beauducel, 2015). In the replication study (study II), participants were trained during the exercise to not blink frequently or avoid blinking during times of interest and epochs with remaining blinks were rejected in the following (Johnson et al., 2008).

Data were segmented into epochs containing the time of interest of the ERPs (see Table 3). Since MFNs typically occur after the response, epochs of MFNs were in all studies time locked to the response. Conversely, in CIT studies, epochs of P300 amplitudes are typically time locked to the stimulus onset and therefore this procedure was also chosen for data of P300s in our CIT studies. In the replication study, all ERPs were analyzed by the same epochs time-locked to the response, as in Johnson et al. (2008). To ensure an ERP neutral baseline, the baseline interval was set to a time before the mean response time for response-locked epochs (about -1,100 ms to -1,000 ms) and to a time before the stimulus for stimulus-locked epochs (-100 ms to 0 ms). Large artifacts in the epochs, such as muscle artifacts or remaining eye blinks, were eliminated by rejecting epochs comprising very high voltages (the exact thresholds are given in Table 3). Thresholds were based on established values for the applied paradigms (Johnson et al., 2008; Olson et al., 2019; Rosenfeld, Labkovsky, et al., 2017; Sai et al., 2020). As outlined in

Chapter 2.3.2, ERPs are relatively small in comparison to the ongoing EEG (Fabiani et al., 2017; Luck, 2014). Therefore, epochs were averaged to optimize the signal-to-noise ratio to make ERPs visible (Fabiani et al., 2017; Luck, 2014). Participants included into data analysis had to have at least about 20 epochs per analyzed categories. Only correct trials were analyzed (meaning responding, as instructed, truthfully or deceptively).

3.2 Quantification of ERP amplitudes

There are two classic methods for quantifying ERP amplitudes: computing peak and mean amplitudes. Peak amplitudes are either the maximal or minimal point of voltage in a pre-defined time window (Donchin & Coles, 1978; Luck, 2014). For mean amplitudes, the average voltage value of a specified time-window is computed (Luck, 2014). Peak amplitudes have a long tradition, especially since they were relatively easy to determine when there were no computer programs available to compute them automatically (Luck, 2014). As Donchin and Coles (1978) pointed out, "It requires nothing but an x-y plot, a ruler, and enough time for this tedious job." After peak amplitudes became standard, they were also applied when amplitudes could be computed automatically (Luck, 2014). However, nowadays, mean amplitudes are frequently used (Luck, 2014). Mean amplitudes better represent the ERP component as a voltage fluctuation extended over time than peak amplitudes (Luck, 2014). Furthermore, mean amplitudes have the advantage that they tend to be less influenced by high-frequency noise than peak amplitudes (Luck, 2014). In deception studies, MFNs are commonly quantified as mean amplitudes (Gibbons et al., 2018; Johnson et al., 2008; Suchotzki et al., 2015). Accordingly, in the research studies of the present thesis, MFNs were quantified as the mean amplitude of a time window occurring shortly after the response at fronto-central electrodes (Johnson et al., 2008). Following Johnson et al. (2008), also P300 amplitudes were quantified as mean amplitudes at a time window around the response at parietal-central electrodes in study II.

Conversely, P300s in CIT studies are usually quantified as peak-to-peak amplitudes at electrode Pz (Gamer & Berti, 2012; Olson et al., 2019; Rosenfeld, Ozsan, et al., 2017). The peak-to-peak amplitude is not, as the name suggests, formed by peaks of single voltage points but by two sliding means. Typically, the most positive mean amplitude of a time segment of 100 ms is searched in a time-window, which is in line

with the usual occurrence of the P300 and its waveform in the grand average of the study. Afterwards, the maximal negative amplitude of a time segment of 100 ms is searched from the midpoint (latency) of the positive peak to the end of the epoch. The difference between the positive and negative amplitude forms the peak-to-peak amplitude (Rosenfeld et al., 2015; Rosenfeld, Ozsan, et al., 2017). It has been argued that the peak-to-peak P300 enables to incorporate more P300 related processes than a single peak amplitude (Soskins et al., 2001). Nevertheless, it is also possible that the peak-to-peak P300 represents additional processes that are not characteristic for P300s (Leue & Beauducel, 2019). Previous research comparing peak-to-peak amplitudes with other ERP quantifications, such as single peaks and mean amplitudes, came to the conclusion that the quantification of the P300 by peak-to-peak amplitudes is superior for detecting concealed knowledge (Leue & Beauducel, 2019; Meijer et al., 2007; Soskins et al., 2001). Accordingly, it has been recommended to compute peak-to-peak P300 amplitudes in CIT studies (Leue & Beauducel, 2019; Meijer et al., 2007; Soskins et al., 2001). Therefore, this approach was also followed in our CIT-studies. All ERP amplitudes were calculated with the statistical programming language R.

3.3 Statistical analyses

Research questions were tested by means of repeated measures ANOVAs. In the deception tasks of the studies, all participants gave deceptive as well as truthful responses. Accordingly, deceptive vs. truthful responses (meaning in the case of CIT studies probe vs. irrelevant items) were entered as a within-subject factor. When Machiavellianism was considered as a possible moderating predictor of the differences in ERPs between deceptive and truthful responses (studies I and II), mean centered Machiavellianism scores were entered as a covariate. When groups of participants with different moral contexts were analyzed (studies I and III), the group variable served as a between-subject factor. Furthermore, the electrode position was entered as a within-subject factor for ERP amplitudes measured at multiple electrodes, which was the case for MFN amplitudes in all studies, and P300 amplitudes in the replication study. In case of violations of the sphericity assumption, degrees of freedoms were corrected by means of Greenhouse-Geisser epsilon. Repeated measures ANOVAS were computed with IBM SPSS.

Table 3

Settings for recording and pre-processing of EEG data

	Study I (CIT study)	**Study II (Replication study)**	**Study III (CIT study)**
EOG	Electrodes at the epicanthis of both eyes, one electrode below the right eye	One electrode below the epicanthus of the left eye, FP1	Electrodes at the epicanthis of both eyes, one electrode below the right eye
Sampling rate	512 Hz	128 Hz	512 Hz
Reference electrode	P9 and P10	pre-auricular sites	P9 and P10
Filtering	P300: 0.16-6 Hz MFN: 0.3-30 Hz	All ERPs: 0.01-35 Hz	P300: 0.3-6 Hz MFN: 0.3-30 Hz
Threshold criteria for rejection of epochs	EEG signal exceeding ±80 μV	EOG signal exceeding ±50 μV during any eight consecutive sampling points	EEG signal exceeding ±80 μV
P300 quantification	Peak-to-Peak amplitude at Pz Search window positive peak: 300-700 ms post-stimulus Search window negative peak: latency of the positive peak until 1300 ms post-stimulus	Mean amplitude at Pz, P3, P4, CP1, CP2 between -100 until 100 ms post-response	Peak-to-Peak amplitude at Pz Search window positive peak: 300-700 ms post-stimulus Search window negative peak: latency of the positive peak until 1300 ms post-stimulus
MFN quantification	Mean amplitude at Fz, FCz, and Cz between 0-70 ms post-response	Mean amplitude at Fz, FC1, FC2, and Cz between 10-80 ms post-response	Mean amplitude at Fz, FCz, and Cz between 0-70 ms post-response
Minimal number of epochs for participants to be included in data analysis	At least 20 artifact free and correct trials of each stimulus of the CIT	At least 16 artifact free and correct trials in one of the analyzed categories	At least 20 artifact free and correct trials of each stimulus of the CIT

Note. Settings were chosen in accordance with the study by Johnson et al. (2008) and previous CIT studies (references are given in the articles of the studies). Further descriptions of the settings can be found in the articles of the studies.

3.4 Participants and deception tasks

Participants were right-handed and had no neurological disorders. They were young adults with a similar mean age of about 22 years (age range: 17-40 years). The gender was equally distributed in all three studies, about 50% were men. Furthermore, in studies with different conditions (study I and III), the gender distribution did not significantly differ between conditions. In all studies, about 40% were psychology students. People who had already participated in one of the studies were not allowed to participate in another study of the thesis. The sample size of all studies was determined with G*Power. In line with previous CIT studies and the study by Johnson et al. (2008), deception tasks were completed on a computer. Participants gave their responses by button presses.

4 Studies

The following chapter gives an overview of the thesis' studies. The design and main findings of the studies will be summarized. Precise details about the methods and results of the studies with their corresponding statistics can be found in the original articles of the studies, which are attached at the end of the thesis. In Table 4 at the end of the chapter, hypotheses of all studies and their results are listed.

4.1 Study I. Individual differences in ERPs during deception

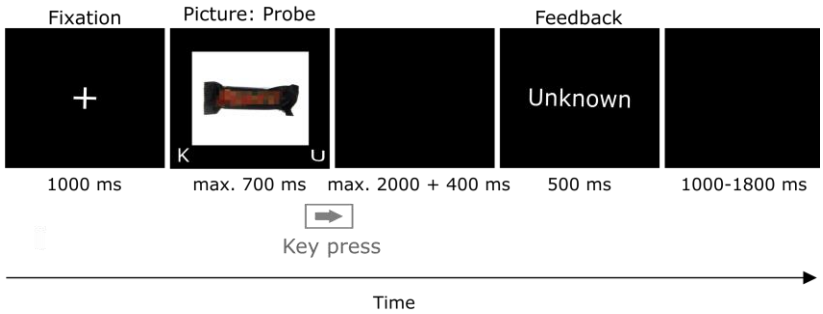
It has already been found that gender and a personality trait associated with moral processing (i.e., sensitivity to injustice) moderate MFN and P300 patterns during the concealment of knowledge (Leue et al., 2012; Leue & Beauducel, 2015). Building upon these findings, an aim of the first research study was to investigate moderating effects of Machiavellianism on MFNs and P300s during a CIT. Moreover, we wanted to take a further step by investigating whether the patterns of P300 and MFN amplitudes differ depending on whether the concealment of knowledge is more likely associated with helping another person vs. saving oneself from trouble. More precisely, the moderation effect of moral involvement on the MFN- and P300-based CIT effect was analyzed (witnessing vs. performing a behavior causing a problem). Since participants got to know the probe items during a social interaction and results of previous studies revealed that women are more inclined to think about relationship aspects in their everyday life and tell more other-oriented lies (Cross & Madson, 1997; DePaulo & Bell, 1996; Erat & Gneezy, 2012; Jaffe & Hyde, 2000; Wark & Krebs, 1996), we additionally investigated moderating effects of gender. Therefore, by considering Machiavellianism, moral involvement, and gender we aimed to get a fuller picture of the determinants of ERP patterns during deception and their interactions.

As outlined in Chapter 2.4, in CIT-studies, participants typically get to know the probe item during an instructed mock-theft and subsequently conceal their knowledge of the probe item during the CIT to cover up that they are guilty of theft. In the present study, participants either got to know the probe item when they performed a behavior leading to a problem (active condition) or when they witnessed another person performing the same behavior (informed condition). Accordingly, in the active condition, participants covered

up a behavior of their own by concealing knowledge of the probe item, whereas in the informed condition participants covered up a behavior of somebody else and this person might hereby avoid potential criticism. Unlike in previous studies, participants were not instructed to perform the behavior causing a problem but performed it spontaneously on their own accord. We intentionally decided for this approach, since otherwise participants in the active condition could possibly blame the instructions for their behavior and may not feel as involved in the behavior causing a problem. During an interaction with an examiner, some participants had the opportunity to take a candy bar, which they either chose to eat (active condition, $n = 31$) or declined (neutral condition, $n = 41$). Other participants witnessed an examiner eating the candy bar (informed condition; $n = 34$). The comparison of the active and informed condition was of primary interest to the study. In the informed and active condition, it was implied afterwards that the candy was not meant to be eaten. Shortly before the start of the CIT, an examiner, who was not present when the candy was eaten, remarked in an angry tone, "Where has all the candy gone?! Now somebody has to go and buy new ones." The other examiner answered with a "Hmm". A rather implicit accusation was given, in order to prevent a very artificial situation and to not offer participants the possibility to excuse themselves. During the CIT, pictures of objects were presented to the participants (see Figure 3 for trial sequence).

Figure 3

Trial sequence of the CIT in study I



Note. Participants indicated for probe items that they are unknown by pressing the right arrow key and the feedback *unknown* occurred. The letters U and K were depicted together with the picture of the item to remind participants that the right arrow key stood for the response *unknown* and the left arrow key for *known*. Adapted from Scheuble and Beauducel (2020).

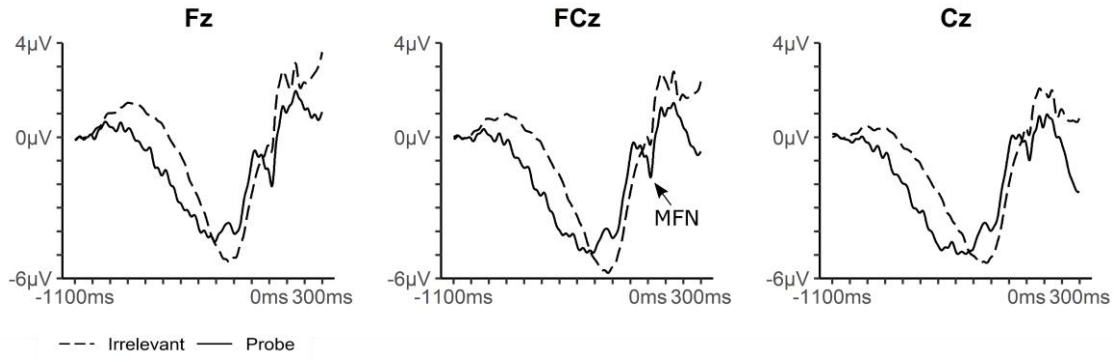
The probe item was the candy bar participants had seen before. Irrelevant items were

formed by similar unknown items. The target item was a similar item that participants got to know before the CIT. Participants were instructed to indicate for the target item that they know it. For all other objects, including probe and irrelevant items, participants were instructed to indicate that they do not know them. Accordingly, they responded honestly for irrelevant items and deceptively for probe items as they actually had seen them before.

After the CIT, participants completed questions about their concerns during lying and a Machiavellianism scale (Henning & Six, 1977). Participants in the active condition indicated to have more concerns during lying than participants in the informed condition. MFN amplitudes were enlarged for probe items, compared to irrelevant items (see Figure 4), revealing that deceptive responses for probe items were accompanied by more conflict than honest responses for irrelevant items.

Figure 4

Response-locked grand averages of fronto-central electrodes



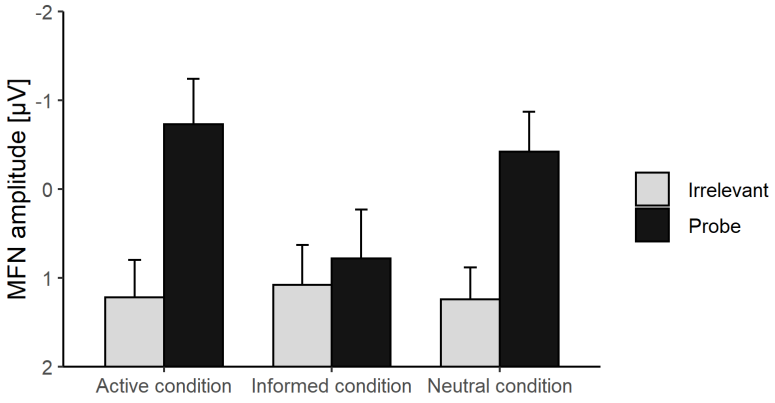
Note. Epochs spanned from 1,100 ms before until 300 ms after the response. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 μV. Adapted from Scheuble and Beauducel (2020).

The MFN-based CIT effect was moderated by moral involvement, revealing that lying as an informed participant was accompanied by less conflict than lying in the active condition (see Figure 5). Furthermore, the effect of moral involvement was moderated by gender: For women but not for men, lying in the informed condition was accompanied by less conflict, as indicated by MFN amplitudes. This is in line with studies finding that women have less scruple to tell other-oriented lies (DePaulo et al., 1996; Erat & Gneezy, 2012; Kashy & DePaulo, 1996). Furthermore, only for women in the informed condition, Machiavellianism moderated the MFN-based CIT effect. For women scoring lower on

Machiavellianism, a larger MFN-based CIT effect occurred, revealing that lying was accompanied by more conflict for them. According to previous research, women have a more interdependent self-construct and, as outlined before, tell more other-oriented lies (Cross & Madson, 1997; DePaulo et al., 1996). When lying has the potential to prevent a social conflict, women high in Machiavellianism could see lying as an appropriate mean to keep their relationships intact. Conversely, women lower in Machiavellianism probably perceive lying in general as morally reprehensible, and it was therefore accompanied by more conflict for them.

Figure 5

MFN amplitudes for the conditions of study I



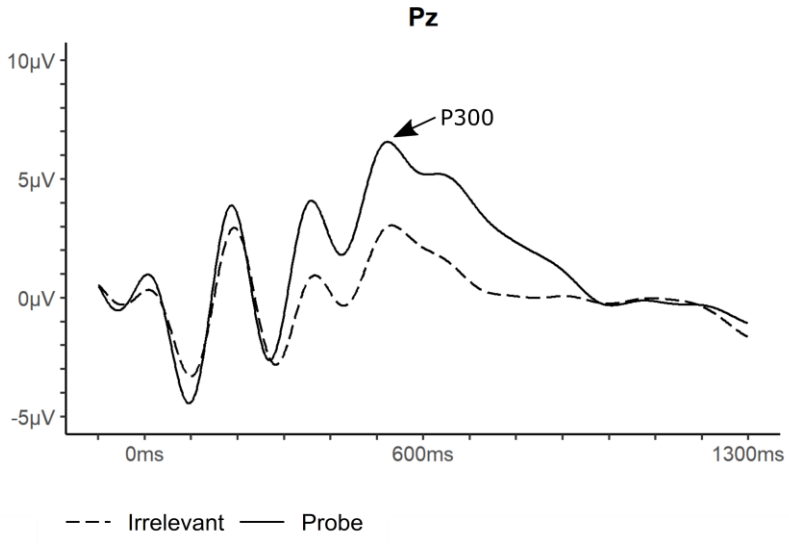
Notes. Error bars indicate standard errors. Active condition= Participants who ate the candy bar, Informed condition= Participants who saw another person eating the candy bar, Neutral condition= Participants who declined the candy bar. A bar chart depicting MFN amplitudes in the conditions for men and women can be found in the article of the study.

In line with previous studies, P300 amplitudes were more positive for probe than irrelevant items, revealing that probe items were more salient (see Figure 6). Furthermore, the P300-based CIT effect (i.e., the difference between probe and irrelevant items) was larger for women than men. For men and women, P300 amplitudes were more positive for probe than irrelevant items, yet for women, probe items known from a social situation were even more salient than for men. This result fits well to the finding of previous studies that women are more likely to consider relationship aspects in their everyday life (Cross & Madson, 1997; Gilligan, 1977; Jaffe & Hyde, 2000; Rothbart et al., 1986; Wark & Krebs,

1996; You et al., 2011), and reveals that women and men may differ in their attention to probe stimuli known from a social context. Unexpectedly, the P300-based CIT effect was not moderated by moral involvement: The difference of P300s between probe and irrelevant items was not larger for participants in the active compared to the informed condition. Accordingly, probe items were not more salient for participants who got to know them while performing a behavior causing a problem, compared to participants witnessing the same behavior. Likewise, Machiavellianism did not moderate the P300-based CIT effect. Overall, the results of the present study highlight the importance of considering individual differences when studying deception. Especially conflicts during deception seem to be different for individuals, as they may also depend on an individual's moral involvement, gender, and Machiavellianism score.

Figure 6

Stimulus-locked grand average at electrode Pz in study I



Note. Epochs spanned from 100 ms before until 1,300 ms after the stimulus. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 µV. Adapted from Scheuble and Beauducel (2020b).

4.2 Study II. Cognitive processes during deception about attitudes revisited

Even though the first study gave important insights about individual differences during deception, it has to be noted that only a certain type of deception was investigated: the concealment of knowledge. To get to know more about general cognitive processes during deception tasks and their determinants, it seemed crucial to additionally study deception in a non-recognition context. Deception studies that do not involve recognition tasks are scarce and there are very few studies that analyze MFNs and P300s (Johnson et al., 2008; Leue & Beauducel, 2019; Suchotzki et al., 2015).

Johnson and colleagues (2008) analyzed lies about attitudes and found different patterns between lies and honest responses of MFNs, P300s[3], and a new ERP they named pre-response positivities (PRP). Even though the results of the study by Johnson et al. (2008) were promising, they were based on a small sample size (N = 17). Furthermore, previous studies investigating deception in non-recognition tasks, found mixed, and also converse results to the study by Johnson et al. (2008). Whereas Kireev et al. (2008) found enlarged amplitudes of an ERP component similar to the MFN during deception, in another study, enlarged MFN amplitudes occurred for honest compared to deceptive answers (Suchotzki et al., 2015). For these reasons, and since the need for replication studies has generally been emphasized, the second study of this thesis aimed to replicate the findings by Johnson et al. (2008). An additional goal of the study was to investigate moderating effects of Machiavellianism on ERP patterns during deception about attitudes.

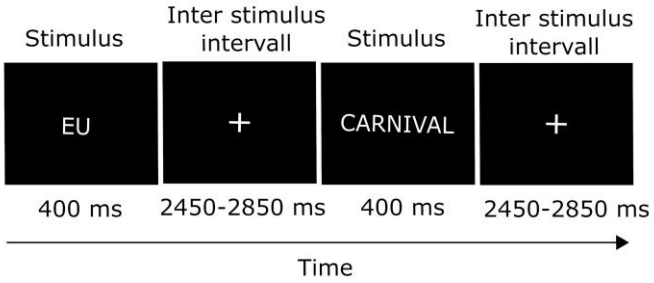
The study was pre-registered prior to data collection (link to pre-registration: https://osf.io/f6w97). Before the deception tasks, participant's attitudes towards religious, political and moral themes, well-known people and preferences were assessed using a questionnaire. During the deception task, participants indicated whether they agree or disagree with the items from the attitude questionnaire (see Figure 7 for trial sequence) while their EEG was recorded. They were instructed to respond truthfully in one block of

[3] In Johnson et al. (2008) and therefore also in our underlying article, the P300 is named LPC. Yet, Johnson et al. (2008, p. 470) ensures that he is referring by the name LPC to the same component as the P300 by writing "the amplitude of the late positive component (LPC) of the ERP (also known as P300)". Likewise, other (deception) researchers use the name P300 and LPC interchangeably (Leue & Beauducel, 2019; Meijer et al., 2007; Polich, 2007). Since in deception studies the name P300 is more common and this name is also used in study I and III, I use the name P300 throughout the thesis.

the task and lie in another block of the task. After the deception task, participants completed a German Machiavellianism questionnaire based on the scale by Christie and Geis (1970) as well as a newer Machiavellianism scale (Henning & Six, 1977; Jones & Paulhus, 2014). Overall, the EEG of 99 participants was recorded.

Figure 7

Sequence of two trials of the deception task in study II



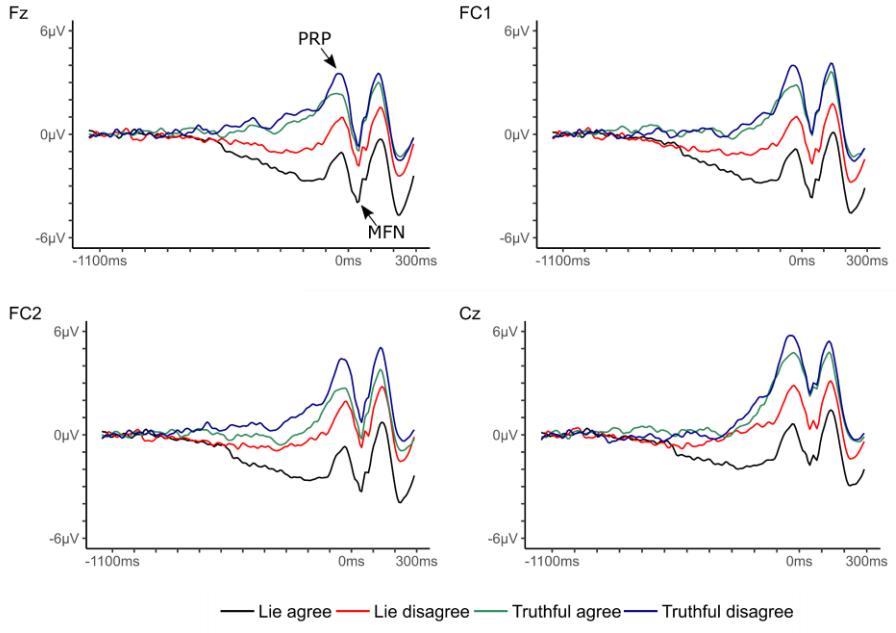
Note. Figure is reproduced from Scheuble and Beauducel (2020a).

Additional to conventional repeated measures ANOVAs, Bayes factors were computed. In contrast to frequentist statistics, Bayes factors also allow for an interpretation of null results. They indicate whether data are more likely under the alternative or null hypothesis. For the following reported effects, results of significance tests and Bayes factors converged.

Results of Johnson et al. (2008) for MFN, P300 and PRP amplitudes could be replicated. Enlarged MFN amplitudes for lies, compared to honest answers, revealed that lies were accompanied by more conflict (see Figure 8). Likewise, suppressed P300 amplitudes for lies, compared to honest responses, indicated that lying relied on additional cognitive resources (see Figure 9). PRP amplitudes were suppressed for lies, compared to honest responses. Suppressed PRP amplitudes have been associated by Johnson and colleagues with strategical monitoring, meaning monitoring processes ensuring that a long-term goal is followed (Johnson et al., 2004, 2005, 2008). Accordingly, lying was accompanied by strategical monitoring.

Figure 8

Response-locked grand averages of fronto-central electrodes in study II



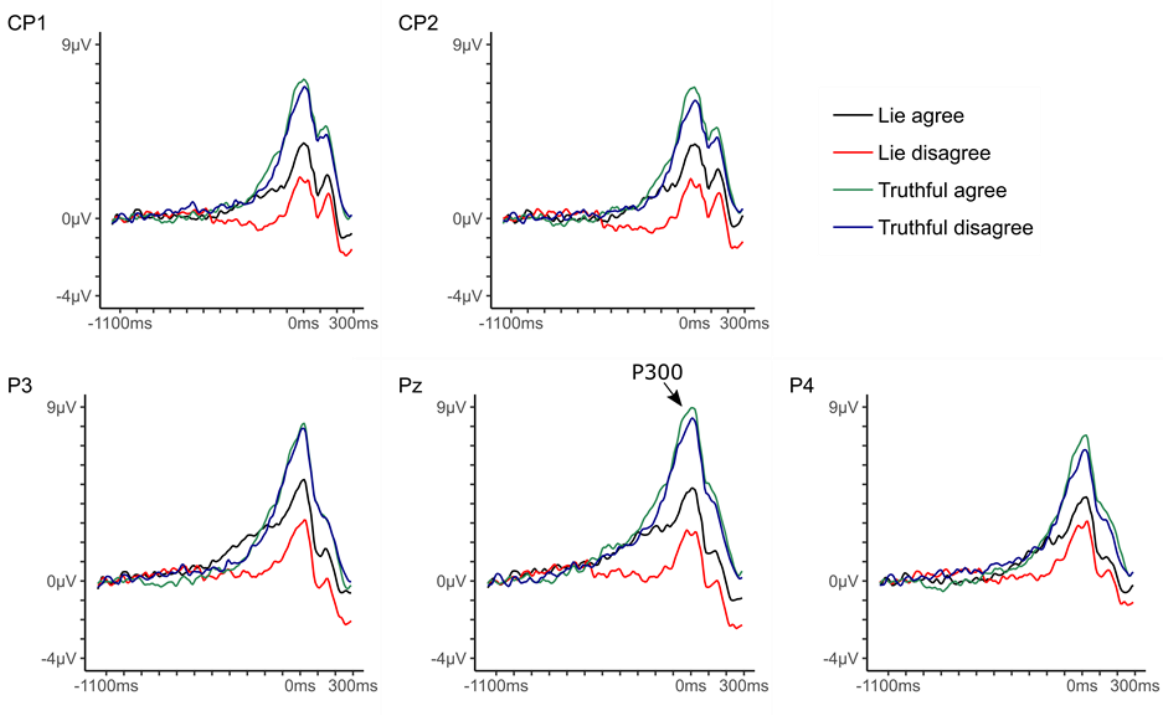
Notes. Epochs spanned from 1,150 ms before to 300 ms after the response. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 µV. Adapted from Scheuble and Beauducel (2020a).

In line with Johnson et al. (2008), we also found moderating effects of the attitudes' valence on ERP amplitudes: Patterns of P300 and MFN amplitudes revealed that lying about positively valued (agree) items were accompanied by stronger conflicts and relied on a greater mental workload than lying about negatively valued (disagree) items. Johnson et al. (2008) proposed that lying about positively valued topics was cognitively more challenging, since it could be perceived as a denial of the self. Conversely, lying about negatively valued items could be perceived as a form of compliance. Exaggerating positive aspects of a negatively valued theme is probably often perceived as socially more acceptable than being honest. This explanation is in line with studies finding that people overestimate the population consensus to a greater extent for items they like than for items they dislike and that it is easier for people to recall positive aspects of negatively judged items than the other way around (Gershoff et al., 2008). Furthermore, responses to positively valued items triggered strategical monitoring, as indicated by suppressed PRP amplitudes for responses to positively valued compared to negatively valued items.

Machiavellianism did not moderate the difference between lies and honest responses for any of the analyzed ERPs. In sum, patterns of P300, MFN and PRP amplitudes proved to be reliable indicators of the cognitive processes during lies about attitudes. Executive processes, like monitoring response conflicts, handling a greater mental workload, and strategical monitoring, seem to be essential when lying about attitudes for both individuals higher and lower in Machiavellianism.

Figure 9

Response-locked grand averages of parietal-central electrodes in study II



Notes. Epochs spanned from 1,150 ms before to 300 ms after the response. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 µV. Adapted from Scheuble and Beauducel (2020a).

4.3 Study III. The P300 and MFN as indicators of concealed knowledge in situations with negative and positive moral valence

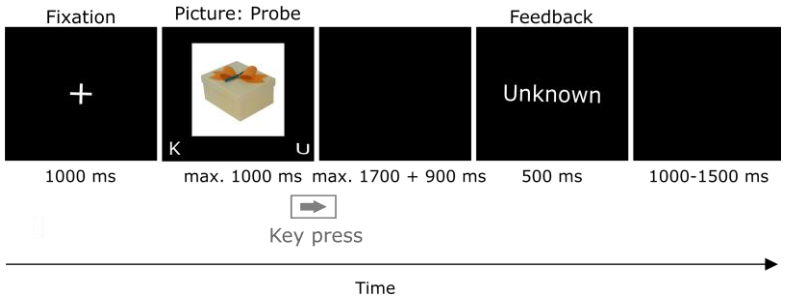
The first study revealed that the MFN-based CIT effect can be sensitive to the moral involvement of the examinee. Furthermore, the results of the second study suggest that the valence of the items of the deception task affect P300 and MFN amplitudes. Even though the involvement of participants differed in the conditions of study I, participants in both conditions got to know the probe items in a negative context. In both conditions, the missing candy caused a problem. Likewise, in existing CIT studies, the scenario before the CIT has usually a morally negative connotation, in that a theft is committed. In this line of reasoning, probe items are in many studies referred to as crime related items (Gamer & Berti, 2012; Rosenfeld et al., 2013; Sai et al., 2014). Hence, it remains possible that the MFN- and P300-based CIT effect also relies on the negative context in which the probe item is typically seen. Therefore, the aim of the third study of the thesis was to investigate whether the moral valence of the situation in which participants got to know the probe item moderates the MFN- or P300-based CIT effect.

For this purpose, some participants (n = 33) committed a classic mock-crime. They stole an object, more precisely a candy box, from an office. A theft harms another person and is therefore considered a behavior with negative moral valence. Other participants (n = 28) performed a similar task that differed in its moral valence. They gave somebody a present by putting a candy box into an office. Since giving a present is usually meant to promote the happiness of others, it formed the behavior with positive moral valence. Besides giving or taking the candy box, the two conditions were kept the same. All objects and the office in the two conditions were identical. Likewise, all following instructions were the same in both conditions. While participants took the CIT, the EEG was recorded. Participants were instructed to indicate whether they recognize the presented stimuli of the CIT (see Figure 10 for trial sequence of the CIT). The probe items consisted of the candy box participants put respectively took from the office and a key that they also saw in the office. We used two probe items and the CIT therefore also consisted of two parts, with keys and gift boxes as picture stimuli, in order to ensure that the results are stable for different stimuli. The irrelevant items consisted of similar boxes

and keys. The target items were formed by a box and a key, with which participants got in touch shortly before the EEG examination. For target items, participants were instructed to indicate that they do know them. For all other items (including probe and irrelevant items), they should indicate that they do not know them. Accordingly, they responded honestly for irrelevant items and deceptively for probe items.

Figure 10

Trial sequence of the CIT in study III

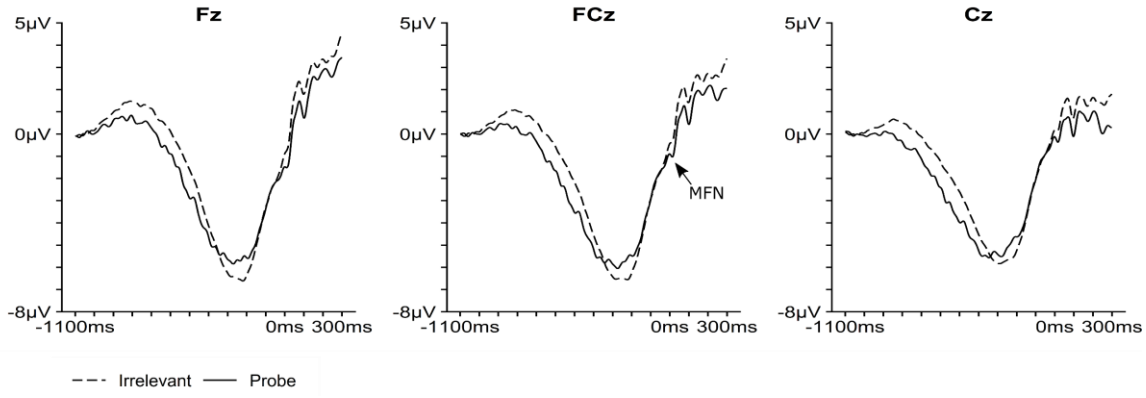


Note. K= known, U= unknown. Adapted from Scheuble et al. (2021).

Additional to repeated measures ANOVAs, Bayes factors were computed. The results of the two analyses converged for the following reported results, unless otherwise stated. In line with previous studies of the thesis, MFN amplitudes were enlarged for probe compared to irrelevant items, indicating that deceptive responses for probe items were accompanied by more intense conflict than honest responses for irrelevant items (see Figure 11). Unexpectedly, the repeated measures ANOVAs revealed that the MFN-based CIT effect was not moderated by moral valence. The Bayes factor was inconclusive and did therefore provide no clear evidence for or against the moderation effect of moral valence. Bayes factors and repeated measures ANOVAs yielded evidence for a CIT effect in both the moral negative and the moral positive condition: When participants got to know the probe items during a behavior with positive moral valence and when they got to know them during a behavior with negative moral valence, MFN amplitudes were enlarged for probe in comparison to irrelevant items. Accordingly, in both conditions, deceptive responses for probe items were accompanied by more response conflict than honest responses for irrelevant items.

Figure 11

Response-locked grand averages of fronto-central electrodes in study III

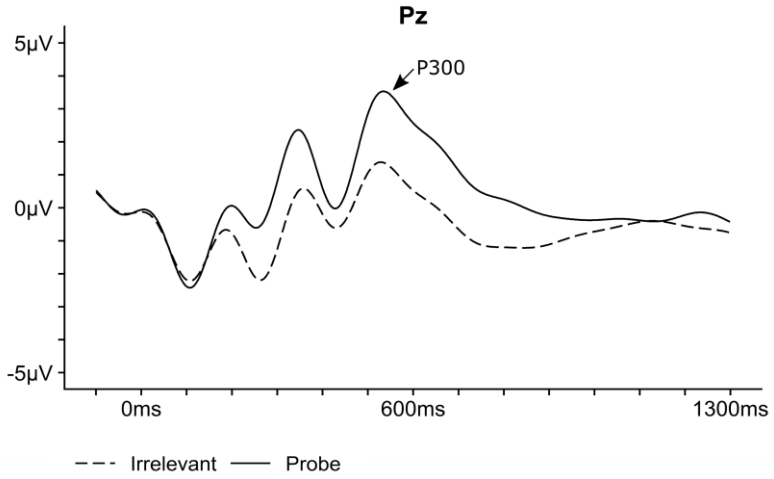


Notes. Epochs spanned from 1,100 ms before until 300 ms after the response. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 µV.

P300 amplitudes were enlarged for probe compared to irrelevant items, revealing that probe items were more salient than irrelevant items (see Figure 12). Moral valence did not moderate the P300-based CIT effect. In the conditions with negative moral valence and positive moral valence, P300 amplitudes were enlarged for probe compared to irrelevant items. Overall, the MFN- and P300-based CIT effect proved to be stable for both a moral negative and positive context.

Figure 12

Stimulus-locked grand average at electrode Pz in study III



Note. Epochs spanned from 100 ms before until 1,300 ms after the stimulus. One tick at the x-axis stands for 100 ms and one tick at the y-axis for 1 µV.

Since in study I interaction effects of gender with moral involvement occurred, one could wonder whether a moderation effect of moral valence occurs in the present study when gender is considered as a between-subject factor. For a better comparability of the results of the two studies in the thesis, I performed additional analyses considering moderating effects of gender. Neither for MFN nor for P300 amplitudes gender moderated the CIT effect in general or in the two conditions. Also, when considering gender as a moderating variable, categorization in significant and non-significant results were the same as reported before. The results of study III and the other two studies will be interpreted as a whole in the following discussion section.

Table 4

Overview of hypotheses and results of the studies of the thesis

	Study I	**Study II**	**Study III**
Deception task	CIT	Deception about attitudes	CIT
Enlarged MFNs for deceptive compared to honest responses	+	+	+
Differences of P300s between items requiring a deceptive and items requiring an honest response	+	+	+
Moderating effects of variables related to the moral context	Moral involvement (witnessing vs. demonstrating behavior leading to a social conflict) moderates the MFN-based CIT effect + Moral involvement moderates the P300-based CIT effect - Exploration of the moderating effect of Machiavellianism on ERPs: A moderating effect of Machiavellianism on the MFN-based CIT effect occurred for female witnesses	Machiavellianism moderates the difference of MFNs between deceptive and honest responses - Machiavellianism moderates the difference of P300s between deceptive and honest responses -	Moral valence (performing a positive vs. negative behavior before the CIT) moderates the MFN-based CIT effect - P300-based CIT effect occurs for both positive and negative behavior (based on the results of study I no moderation effect was expected) +

	Study I	Study II	Study III
Additional hypotheses resulting from the specific study design/topic:	Moderating effect of moral involvement on MFN-based CIT effect is more pronounced for women than men +	Larger MFNs and suppressed P300s for deception about positively valued items, compared to deception about negatively valued items +	
	Difference of P300s between probe items seen in a social context and unknown irrelevant items is larger for women than men +	Attenuated PRPs for deceptive compared to honest responses +	
		Attenuated PRPs for positively valued compared to negatively valued items +	

Notes. + = Results were in line with hypothesis, - = Hypothesis could not be confirmed

5 Discussion

The goal of the thesis was to investigate cognitive processes during deception tasks, as indicated by P300 and MFN amplitudes. Thereby, I aimed to reveal possible determinants of the cognitive processes during deception with a special focus on variables related to the moral context, that is, moral involvement, Machiavellianism, and moral valence. For ERP-based deception studies, certain designs are regularly applied, for instance, they often utilize CITs with a prior instructed mock-theft (Gamer & Berti, 2012; Lui & Rosenfeld, 2008; Sai et al., 2020; Winograd & Rosenfeld, 2014). In the deception studies of the present thesis, we investigated whether the different patterns of MFN and P300 amplitudes can also be repeatedly found when deviating from this design. The deception tasks of the studies of the thesis all involved non-forensic scenarios. Furthermore, in study I, the scenario before the CIT was not instructed, in study II, a deception paradigm other than the CIT was applied, and in study III, some participants demonstrated a pro-social behavior before the CIT. Altogether, the studies of the thesis aimed to underpin the theoretical basis of deception tasks by investigating on which situations the differing patterns of MFN and P300 amplitudes in deception task can be generalized and by which variables they are moderated. The results of the studies will be discussed in the following. Patterns of MFN amplitudes in deception tasks and their sensitivity to variables related to the moral context will be interpreted. Since the P300 proved to have a dual-nature in the different deception paradigms, the results of P300 amplitudes will be interpreted first for CITs and afterwards for deception about attitudes. When certain results have direct implications for future research, they will be outlined after the interpretation of the results. Further outlooks will be given in a separate section of the discussion.

5.1 The MFN as an indicator of conflicts during deception and the moderating role of the moral context

In all studies of the thesis, MFN amplitudes were enlarged for deceptive compared to honest responses. Accordingly, deception was in general accompanied by more conflict than being honest. This pattern of MFN amplitudes was found across different deception

tasks, that is, for CITs (study I and III) and for deception about attitudes (study II). Therefore, the MFN seems to be a promising marker of conflicts involved in deception.

Nevertheless, some variables related to the moral context of the deception task moderated the patterns of MFN amplitudes. This became apparent in study I. The MFN-based CIT effect was attenuated for participants witnessing another person performing a behavior causing a problem (informed participants), compared to participants who performed the same behavior (active participants). Deception was accompanied by less conflict for informed compared to active participants, and this effect occurred especially for women. In this line, we also expected in study III that the moral valence of the pre-CIT scenario moderates the MFN-based CIT effect. Participants either saw probe items while performing a behavior with positive moral valence (giving a present) or while performing a behavior with negative moral valence (committing theft). Subsequently, they responded deceptively during the CIT that they do not know the probe item. We did not find a moderation effect of moral valence. The pattern of MFN amplitudes revealed that participants performing a behavior with positive moral valence did not experience significantly less intense conflicts during deception than participants performing a behavior with negative moral valence.

Altogether, the results of the two studies reveal an important difference between moderating effects of moral involvement and valence on the conflicts during deception. The results of study I suggest that moral involvement can moderate the conflicts during deception. Possibly, concealing knowledge of the probe item as an informed participant was linked to covering up a behavior of somebody else. Since it could therefore be considered as lying for somebody else, it was accompanied by less guilt and conflict. Conversely, the results of study III reveal that the conflicts during deception are stable for situations with positive vs. negative moral valence. Even though participants demonstrated a positive behavior in the moral positive condition of study III, they still covered up their own behavior by concealing knowledge of the probe item and therefore lying could possibly not as easily be perceived as prosocial and accompanied by less conflict. However, this explanation is tentative and needs further support from future empirical data. The investigation of MFNs in deception tasks in general and in particular of moderating effects of moral variables on MFNs is still in its infancy. Especially against this background, future studies are required to test whether the results regarding the

moderating effects of moral involvement as well as moral valence are replicable. In this regard, it also has to be noted that the moderating effect of moral valence was not significant and the significance test and Bayes factors yielded evidence for an MFN-based CIT effect in both the moral negative and moral positive condition of study III. Yet, the Bayes factor of the moderating effect of moral valence was inconclusive. Future replication studies and studies comparing directly the effect of moral valence and moral involvement within one study could give further evidence whether moral involvement has to be established as a moderating variable of the MFN-based CIT effect and whether moral valence probably does not moderate the MFN-based CIT effect in general.

5.2 Moderating effects of Machiavellianism on MFNs

Besides elucidating moderating effects of moral valence and moral involvement on MFN amplitudes, a further aim of the present thesis was to investigate moderating effects of individual differences in Machiavellianism. We did not find a general moderation effect of Machiavellianism for the difference of MFN amplitudes between honest and deceptive responses. Accordingly, conflicts during deception were not in general reduced for individuals higher in Machiavellianism. Instead, Machiavellianism moderated the conflicts during deception only under certain conditions. A moderation effect of Machiavellianism did not occur for deception about attitudes (study II). Yet, in study I, Machiavellianism moderated the conflicts during deception for women in the informed condition. For women higher in Machiavellianism the MFN-based CIT effect was less pronounced. Accordingly, they experienced fewer conflicts when they lied.

A possible explanation for the result that the moderation effect of Machiavellianism occurred under these conditions is the following. One characteristic of Machiavellianism is the care about one's own reputation, as well as reputation building (see Chapter 2.5.2). Previous research found that women have a more interdependent self-construal than men (Cross & Madson, 1997). Likewise, women tend to tell more other-oriented lies (DePaulo et al., 1996). As outlined before, informed participants rather covered up the behavior of another person than their own by responding deceptively during the CIT. Thereby, the other person could also possibly avoid further criticism. In this situation, women high in Machiavellianism may perceive lying as an acceptable mean to avoid an argument and to care for their relationships as well as social reputation. In contrast,

women lower in Machiavellianism possibly perceive lying in general as ethically objectionable and hence experience more conflicts when responding deceptively.

That a moderation effect of Machiavellianism was only found under certain conditions, is in line with a current theory by Jones and Mueller (2021). They reviewed research findings of Machiavellianism and summarized that many effects of Machiavellianism are context dependent. Instead of asking the research question what antisocial behavior can be predicted by Machiavellianism, they concluded that it is more reasonable to ask when Machiavellianism predicts antisocial behavior. Keeping the guiding principle of Machiavellianism "the end justifies the means" in mind, a research approach in which the context and possible interactions with the situation are considered, seems to be compatible with the conceptualization of Machiavellianism. Considering personality characteristics in combination with the situation is in tradition with interactionism theories. Already, Hartshorne and May (1928) found that the correlation of deceitful behavior between different situations is low and therefore highlighted the importance to consider the specificity of the situation when predicting deceitful behavior. Jones and Mueller (2021) proposed characteristics of trait-relevant situations for Machiavellianism. They outlined that Machiavellianism predicts antisocial behavior when the rewards of the antisocial behavior are higher than the risks out of it. It should be noted that in the informed condition of study I, the rewards of deception were also probably higher (prevent a possible argument and keeping relationships intact) than its risks. The finding that Machiavellianism does not moderate conflicts during deception in general but only in certain situations suggests that the trait relevance of the situation may also be decisive for effects of Machiavellianism on cognition. Therefore, it might be worthwhile to focus in future studies further on the question in which situations Machiavellianism moderates conflicts during deception.

5.3 P300s as stable indicators of the salience of probe items in CITs

Beyond MFN amplitudes, we investigated in all studies the P300, an ERP component that is regularly analyzed in deception paradigms, especially CITs. The P300 proved to be a stable indicator of increased salience of probe items requiring a deceptive response,

compared to irrelevant items requiring an honest response. In the two studies applying a CIT (study I and III), P300 amplitudes were enlarged for probe compared to irrelevant items. Getting to know the probe item in a forensic context does not seem to be a prerequisite for the P300-based CIT effect. In both studies, the P300-based CIT effect occurred in non-forensic situations. Furthermore, in all studies, P300 amplitudes were not moderated by variables related to the moral context of the deception tasks. Neither Machiavellianism, nor a prosocial vs. antisocial framing of the deception scenario moderated the P300-based CIT effect. The results of study I revealed that probe items were not less salient for individuals higher in Machiavellianism. In other words, probe items did not appear less meaningful for individuals who tend to have less scruple and ethical concerns to lie. Moreover, neither moral involvement (study I) nor moral valence (study III) moderated the P300-based CIT effect.

The result of study I that moral involvement did not moderate the P300-based CIT effect was unexpected. In a former study by Rosenfeld, Ozsan, et al. (2017), a greater P300-based CIT effect was found for perpetrators than witnesses of a mock-theft. Similarly, we expected probe items to be more salient for participants who performed a behavior causing a problem than for participants who witnessed somebody else demonstrating the same behavior. However, the P300-based CIT effect did not significantly differ between the two conditions. Yet, this finding is in line with results from some other studies: Jang et al. (2013) also found no difference of the P300-based CIT effect between witnesses and perpetrators of a mock-crime. Furthermore, in a study by Winograd and Rosenfeld (2014), the P300-based CIT effect did not differ between perpetrators of a mock-crime and participants who were only informed about the crime but did not commit it. A possible explanation for the differing results is that in the study in which a moderation effect was found (Rosenfeld, Ozsan et al., 2017), witnesses and perpetrators got different instructions for the CIT: Perpetrators were instructed to conceal their knowledge of the probe item, whereas witnesses were instructed to try to reveal their knowledge of the probe item. Conversely, in the studies in which no moderation effect was found (study I of the present thesis; Jang et al., 2013; Winograd & Rosenfeld, 2014), participants of the different conditions got the same instructions during the CIT. All of them were instructed to conceal their knowledge of the probe items. Heightened deception awareness as well as instructing participants to respond deceptively during the CIT can increase the

P300-based CIT effect (Rosenfeld et al., 2012; Verschuere et al., 2009). Accordingly, the difference in the instructions of the CIT (trying to conceal vs. reveal the knowledge of the probe item as witnesses/informed participants) can offer a possible explanation for the differing results between study I and the study by Rosenfeld, Ozsan et al. (2017).

Moreover, in study III, we found that the P300-based CIT effect also occurs for participants who did not perform a behavior causing harm to others. For both, participants who got to know the probe item while performing a behavior with positive moral valence (giving a present) and participants who performed a behavior with negative moral valence (committing theft), probe items were more salient than irrelevant items during the CIT. It should be noted that also in study III, participants of both conditions were instructed to conceal their knowledge of the probe items. Altogether, the results of study I and III imply that neither moral valence nor moral involvement determine the P300-based CIT effect. The greater salience of the probe item does not seem to depend on getting to know the probe item while performing a negative, crime-related behavior. Probe items did not appear to be more salient or meaningful when they were related to a negative compared to a positive behavior. Likewise, the results of study I imply that participants do not have to touch the probe item themselves or be actively involved in the relevant behavior during which they got to know the probe item for the probe to be more salient than irrelevant items. Also, for informed participants, probe items were more salient than irrelevant items. The results seem to indicate that moral valence and moral involvement do not represent confounding variables of the P300-based CIT effect, hindering its validity. They suggest that the P300-based CIT may also be applicable to reveal concealed knowledge about morally positive connoted situations, as well as concealed knowledge of informed people. Whereas MFN amplitudes indicating conflicts during deception seemed to be sensitive to some variables related to the moral context of the CIT, the salience of the probe items, as indicated by P300 amplitudes, proved to be stable for all analyzed moral contexts. It seems like the salience of the probe item, which is based on basic processes as the recognition of an item and its additional concealment, is unaffected by the moral context in the form of moral involvement and valence.

5.4 Patterns of P300s for deception about attitudes

In a deception task differing from the CIT, a converse pattern of P300 amplitudes occurred (study II): When participants lied about their attitudes, P300 amplitudes were suppressed for deceptive compared to honest responses. This pattern of P300 amplitudes was expected and replicated the findings of Johnson et al. (2008). The dual nature of the pattern of P300 amplitudes in CITs and deception task that do not rely on the recognition of an item has to be considered in deception tasks. In CITs, probe items are distinct from irrelevant items for people who recognize the probe items and accordingly conceal knowledge about them. Their attention is therefore drawn to the known probe items. They appear more significant than the irrelevant items, which goes along with enlarged P300 amplitudes. Conversely, in deception tasks beyond the CIT, in which participants give deceptive and honest responses with the same frequency and the stimulus material does not consist of known vs. unknown items, the mental workload hypothesis more likely holds: Deception requires more cognitive resources and can be considered an additional task. The attention is therefore drawn away from the stimulus, which goes along with suppressed P300 amplitudes for deceptive compared to honest responses. This pattern of P300 amplitudes could repeatedly be found for deception about attitudes; in the study by Johnson et al. (2008) in a US sample and in our study for a German sample. Therefore, the P300 seems to be a promising indicator of the mental efforts during deception in non-recognition tasks.

Analyzing P300 amplitudes–in addition to MFN amplitudes–in deception tasks seems to have the potential to better get to know the cognitive processes during deception. By analyzing these ERPs not only in CITs but also other deception tasks, ERP patterns and cognitive processes which are specific for certain deception paradigms and those that generally occur during deception can be uncovered. Whereas MFNs indicated stronger response conflicts in both CITs and paradigms involving deception about attitudes, patterns of P300 amplitudes were different in the two deception paradigms. P300s indicated a greater mental workload for deception about attitudes, though in CITs the salience of the probe items overrode the mental workload effect. Nowadays, ERP-based deception studies mainly apply CITs and the investigation of ERPs in other deception paradigms seems to standstill. However, the results of the thesis' studies imply that differing patterns of ERPs can not only be repeatedly found in CITs but also in other deception paradigms,

more precisely for lies about attitudes. The results of the study are encouraging to further investigate P300s, MFNs, and also PRPs in non-recognition deception paradigms. In this regard, it also has to be considered that the CIT has a restricted scope. It can only be applied when the researcher knows a relevant stimulus, of which a certain target person, such as the perpetrator of a crime, and only this person, tries to conceal knowledge about (Farwell, 2012). Investigating ERPs also in non-recognition deception tasks has therefore not only the potential to get to know more about the general cognitive processes during deception, but it can also possibly widen the scope of ERP-based deception detection in the long run. With the MFN and P300, two ERPs are already available that reproducibly indicate stronger conflicts and mental effort of deceptive compared to truthful responses, at least for deception about attitudes. It remains for future studies to further investigate whether MFN and P300s are also reproducible markers for conflicts and mental effort during deception in other non-recognition contexts.

In this context, I want to highlight that deception research can benefit from studies finding different patterns of ERPs between deceptive and honest responses but also from the reporting of findings when no significant difference in ERPs between deception and honest responses occurs. Only this allows to get an overarching view on the mental efforts and conflicts during deception. It is important to know when mental efforts and conflicts are more intense for deceptive compared to honest responses, as well as when and why this is not the case. Especially for a research field that can always be viewed with a focus on its practical application (in this case deception detection), the possibility of a publication bias needs to be prevented. For this purpose, preregistration and registered reports can be helpful (Scheel et al., 2021). An emerging number of journals, also those important for the ERP-based deception research, already introduced these publication forms, and it could further the deception research when journals more regularly offer publication forms like registered reports, and they do not represent exceptions.

As outlined before, together the studies of the thesis also underlined the dual nature of P300s in different deception paradigms. Studies investigating deception based on the mental workload hypothesis and CIT-studies, which are based on the salience hypothesis, are currently two mainly independent research streams. This becomes also apparent in that different methods are predominantly used for studies based on the salience and those based on the workload hypothesis, e.g., for the pre-processing of their data (i.e.,

filter settings, amplitude quantification etc.). The thesis' studies revealed that patterns of P300 amplitudes based on the salience as well as mental workload hypothesis can reproducibly be found. As a next step, it seems worthwhile to conceptually align CITs and other deception paradigms and investigate where the mental workload hypothesis holds and at which point the salience of a stimulus overrides the mental workload effect of deception. For instance, in a task in which participants lie about their attitudes, the frequency of lies and honest responses could be equalized to those in CITs (in which participants lie less frequently for probe items than they give honest responses for irrelevant items). Based on the investigation of such paradigms at the border between CITs and other deception paradigms, it can more thoroughly be analyzed when P300s are enlarged vs. suppressed for items requiring a deceptive response. Furthermore, it could be worthwhile to compare the differing methods for the quantification of the ERPs of the two research streams in future studies. For example, in CIT studies, peak-to-peak P300 amplitudes are predominantly computed (Gamer & Berti, 2012; Olson et al., 2019; Rosenfeld, Ozsan, et al., 2017), whereas in deception paradigms beyond the CIT P300s are usually quantified as mean amplitudes (Dong et al., 2010; Johnson et al., 2008; Meek et al., 2013; Suchotzki et al., 2015). Therefore, in future studies, the quantification of P300 amplitudes as mean and peak-to-peak amplitudes could be compared for both CITs and non-recognition deception paradigms. Soskins et al. (2001) already compared the quantification of P300s as peak-to-peak amplitudes with a single positive sliding mean and found that the quantification with the peak-to-peak amplitude is superior for the detection of concealed knowledge. Yet, it seems additionally important to compare the quantification of the P300 as a peak-to-peak amplitude with the quantification of the P300 as a mean amplitude for deception paradigms that do not involve the concealment of knowledge.

Additional to replicating the main ERP results for deception about attitudes found in the study by Johnson et al. (2008), we also analyzed the moderating effect of Machiavellianism on the ERPs. As in the CIT study (study I), Machiavellianism did not moderate the difference of P300 amplitudes between deceptive and honest responses about attitudes. Together the studies revealed that in both deception paradigms, when P300s indicated a greater salience of probe compared to irrelevant items, and when P300s indicated a greater mental workload for deceptive compared to honest responses about attitudes, individual differences in Machiavellianism did not moderate the patterns of P300

amplitudes. With regard to the results of study II, this meant that deception about attitudes were not cognitively less challenging for individuals higher in Machiavellianism. This result was also in line with the patterns of MFN amplitudes, indicating that conflicts during deception about attitudes were not reduced for individuals higher in Machiavellianism.

5.5 Limitations

The implications of the results hold for the conditions of the respective studies, and a rash generalization to other contexts should be avoided. Therefore, limitations of the studies are outlined in the following.

First, characteristics of the sample of the present studies have to be considered. This seems especially important for the analyses of the moderating effects of Machiavellianism on ERP patterns. The participants of the studies had mainly an academic background, and it was not secured by a pre-selection that the sample comprises individuals with low and middle scores in Machiavellianism as well as many individuals scoring high on Machiavellianism. As described before, we found a moderation effect of Machiavellianism on MFN amplitudes only in a certain condition of one study (i.e., for women in the informed condition in study I). It is possible that results for Machiavellianism are more conclusive or that moderation effects of Machiavellianism more likely occur when individuals with more extreme scores in Machiavellianism are considered. Either way, future studies investigating the moderation effect of Machiavellianism on ERPs during deception for individuals with extreme values in Machiavellianism (e.g., by a pre-selected sample, prisoners or a clinical sample) could complement the findings of the present thesis.

Furthermore, in all three studies, data from young adults were analyzed. When comparing P300 amplitudes over the lifespan, P300 amplitudes are most pronounced for this age group and decrease for older people (Dinteren et al., 2014). Likewise, it has already been found that MFN amplitudes can differ between younger and older adults (Tays et al., 2008). Thus, it is questionable whether the thesis' results can be generalized to older adults, meaning especially people older than 40 years. Considering that ERP-based deception paradigms mainly involve young adults and that there are to the best of my knowledge no P300- or MFN-based deception studies that specifically analyzed data from

older adults, it seems especially important to further investigate whether the found ERP effects also hold true for this population.

Second, it has to be noted that the conditions of the studies did not involve high stake situations. In study I, investigating the moderating effect of moral involvement, missing candy caused a problem. Accordingly, ERP patterns for a small social conflict were elucidated. We chose this approach, since observing a severe form of misbehaving as well as leading participants to demonstrate such behavior could possibly result in a very artificial situation in a laboratory study, which could in turn hinder the moral involvement of participants in the situation. Moreover, the generalizability of the CIT effect to social situations beyond crimes was also an important aim of the present thesis. In accordance, the probe stimuli were objects that are regularly present in experimental settings (i.e., candy). However, this also means that the study can only provide preliminary hints for the cognitive processes and corresponding ERPs for a CIT involving more severe misbehavior. Likewise, in study III, participants either demonstrated a behavior with moderate positive or moderate negative moral valence. At a first step, we found it important to compare a negative behavior that is regularly investigated in CITs (taking an object in a theft) with a positive behavior of similar intensity (giving an object as a present). However, from the results of the study, it cannot be concluded that the same patterns of ERPs appear when comparing behaviors with a very positive vs. very negative moral valence or when the consequences of the behaviors are more severe.

On a related note, participants in study III were, as in previous mock-theft studies, instructed to demonstrate the behaviors with negative and positive moral valence. Despite the fact that participants of the study were repeatedly informed–before the examination and in the letter in which the instructions were given–that they could quit the task any time, it is possible that moderating effects of moral valence on P300 or MFNs are present when people demonstrate a behavior of their own choice. In this context, a study by Nahari et al. (2017) seems important. They investigated the effect of the free will to demonstrate a mock-crime on psychophysiological data (electrodermal and respiration measures) and response times during a CIT. The CIT-effect did not differ between participants choosing on their own to either commit a mock-crime or another task, and participants who were instructed to commit one of the tasks. Yet, it is possible that, especially for effects of moral valence, it is decisive that the behavior is self-initiated.

Finally, it has to be highlighted that I analyzed deception in regularly used deception paradigms of ERP-studies and wanted to underpin their theoretical basis. Yet, the deception paradigms differ in important aspects from deception in everyday life and might not capture all of its features. They are simplified deception tasks under controlled laboratory settings. Therefore, a harsh generalization of the found results to lies in everyday life is premature and should be avoided. For instance, participants of the studies pressed buttons for responding deceptively. In real life, participants often speak when responding deceptively. Speaking is accompanied by muscle activity, for instance of the tongue, which can distort the analyzed ERPs (Luck, 2014). This is especially problematic for response-locked ERPs, like the MFN. In this line, ERP-based deception paradigms, and also those used in the present thesis, mainly involve responses as button presses rather than verbal answers. Furthermore, the social component in the deception paradigms was low. We applied classic CIT paradigms, in which participants respond deceptively during a computer task and only get feedback by screen displays. Likewise, the deception paradigm comprising attitudes was a computer task, as in the replicated study. In real life, participants are possibly more focused to not only control their responses but also their non-verbal expressions during deception. In this context, a study by Carrión et al. (2010) seems worth noting. They investigated data of participants who sat face-to-face to another person while giving deceptive responses by button presses. They analyzed N450s, an ERP that is also closely associated to conflict monitoring, and found increased conflict monitoring for deceptive compared to honest responses. However, for face-to-face deception studies, it has to be carefully avoided that experimenter effects occur. This seems especially important when the moderating effect of multiple conditions or of personality constructs are analyzed, and examiners can possibly differ between conditions or display different non-verbal behavior towards different personalities. Therefore, we also did not modify the CIT and the deception paradigm involving attitudes in our studies. Nevertheless, new technologies, like virtual realities and mobile EEGs, can be promising to investigate in future studies deception in more ecologically valid paradigms. In sum, it is possible that face-to-face lies in everyday life may comprise additional cognitive load or inhibition processes that are not captured in the deception paradigms of the studies of the thesis. Therefore, the results of the study currently only allow interpretations for cognitive processes in the corresponding deception tasks and can only give hints which ERP

patterns might be expected for real-life face-to-face lies. Nevertheless, it also should be noted that this is not crucial for the application of the analyzed deception paradigms in order to detect deception, since they are intended to be applied also in real-life settings in their current form as a computer task.

5.6 Outlook

While the previous sections already included some hints for future work, in this section a more generic outlook on the field is given. In the thesis' studies, we departed from classic scenarios of deception paradigms and analyzed deception in non-forensic situations. To ensure that findings of deception studies are generalizable to multiple situations, further studies are needed analyzing deception in other than mock-theft scenarios.

For instance, the ERP-based deception research could profit from studies investigating deception at the workplace. When thinking about personnel selection or misconduct at work, it becomes apparent that deception is also present at the workplace. However, there are, to the best of my knowledge, no ERP studies investigating deception in a work-related context. Therefore, investigating deception in these situations could open up a possible application setting for ERP-based deception detection and further the understanding of deception taking place at the workplace. Also, in the workplace, the moral background of deception can vary. For instance, employees can lie for their own personal gain or for the benefit of the organization (Sims, 2002). Analyzing deception with these different goals could complement our findings regarding the moderating effect of pro- vs. antisocial settings on the processes during deception.

In a similar vein, in the forensic context many crime situations are not yet investigated, and future studies should involve more diverse crime scenarios. ERP-based deception studies typically investigate thefts, e.g., of a ring. However, previous cases in which the application of ERP-based CITs have been considered in court, were predominantly murders (Farwell, 2012). Therefore, it is important that the ERP-based CIT-effect is generalizable to multiple situations and also its specific applications. A further promising application of ERP-based CITs, which is not yet well investigated, are planned crimes. Applying the CIT to planned crimes has the potential to prevent crimes from happening. Since in our study a P300-based CIT effect also occurred for participants who did not actually get in touch with the probe item but only knew it, our results may suggest

that the CIT effect also holds true for crimes that are planned but not (yet) committed. Indeed, Meixner and Rosenfeld (2011) found a P300-based CIT effect for a planned terrorist attack. It seems interesting to get to know whether the MFN-based CIT effect is also generalizable to planned crimes. Recall our findings in study I that the MFN-based CIT effect was similar in the active and neutral condition (in which participants knew about the probe item but did not cause a problem) but differed between the active and informed condition. These findings may suggest that an MFN-based CIT effect occurs for planned crimes as long as the participants themselves planned to perform them.

The result that the P300-based CIT effect also occurred for participants who only knew the probe item but did not perform the problematic behavior also suggests that informed participants or witnesses cannot be differentiated from people who actually misbehaved while getting to know the probe items. Accordingly, a different method seems to be needed to differentiate these two groups of people. The guilty action test (GAT), an adaption of the CIT, has been introduced for this purpose in the context of polygraph measures (Bradley et al., 1996). The GAT also involves probe and irrelevant items, but instead of asking examinees whether they recognize the probe item, they are directly asked whether they committed the behavior in question, e.g., “Did you steal the ring?”. Both guilty examinees and witnesses, should recognize the probe item. However, guilty participants lie when denying these questions, whereas witnesses tell the truth. It has been questioned that polygraph tests involving the GAT are successful in differentiating between guilty and informed people (Gamer, 2010). Therefore, it seems worthwhile to investigate whether the GAT based on P300s and MFNs is able to differentiate between guilty, informed people and people who do not know the probe item, by indexing both the salience of the probe item and the conflicts during deception.

P300 amplitudes proved to be stable across different manipulations of the moral context in the present studies. Yet, gender moderated the P300-based CIT effect when participants saw the probe item during a social interaction. We explained this result with women having a greater focus on social stimuli. It is possible that moral variables that are more strongly related to the attentional focus, moderate P300 amplitudes. For instance, it would be interesting to investigate moderating effects of xenophobia. When a participant has to conceal knowledge about the identity of a searched person, this person could share relevant characteristics (e.g., ethnicity, gender) with the examinee or not. Based on

previous research, competing expectations of P300 patterns seem possible for this comparison: The P300-based CIT effect proved to be more pronounced for personally more relevant items (Rosenfeld et al., 2006). When the examinees conceal knowledge about a person who shares relevant features with themselves, this person could be perceived as personally more relevant. Accordingly, an enlarged P300-based CIT effect would be expected when the examinee is more similar to the searched person. Conversely, in categorization tasks, enlarged P300 amplitudes were found for faces of participants ethnical outcompared to their ingroup (Dickter & Bartholow, 2007; Pesciarelli et al., 2021). Based on these findings, an enlarged CIT effect would be expected when the searched person differs from the examinee, for instance by having a different ethnicity.

Finally, it should be noted that even though the P300 and MFN are promising for revealing deception, they both have their limitations. There is, at least not yet, a Pinocchio's nose indicating deception in all circumstances. The MFN can also be sensitive to the moral context of deception. The P300 seems to be a stable indicator of concealed knowledge in CITs. However, the concealment of knowledge is just one form of deception. For deception paradigms beyond the CIT, converse P300 patterns occur and P300s in non-recognition deception paradigms are also investigated less. As outlined in the introduction, many methods are available for the detection of deception, and they all seem to have their limitations. The different methods for detecting deception are usually investigated separately in different research studies, different journals and also often treated as competing methods for the detection of deception. Yet, a multi-method approach comprising, for instance, ERPs, fMRI, verbal and non-verbal signs for the detecting of deception could help to overcome the specific limitations of the single methods. Investigating converse results of these different deception measures could help to further understand their determinants and the processes involved in deception. Likewise, different paradigms can be applied to find the best way to combine their results to increase deception detection rates. For instance, the implicit association test (IAT) is another paradigm that can be used to uncover concealed knowledge or attitudes (Greenwald et al., 2009; Sartori et al., 2008). Hu and Rosenfeld (2012) found that a combination of the results of an IAT and CIT outperforms any single measure. It seems promising to further examine the best way to combine different paradigms to detect deception. Nowadays, conducting studies across different research groups has been promoted (Wacker, 2017). Even though a multi-method

approach can be time-consuming and the expertise of different researchers is needed, this could be worthwhile to find an optimal way for detecting deception.

5.7 Conclusion

In all studies of the thesis, involving various deception paradigms, MFN as well as P300 amplitudes differed for deceptive compared to honest responses. Previous deception studies rarely investigated MFN amplitudes. The thesis' studies thus contribute to establishing the MFN as an ERP component that is worthwhile to investigate in deception studies additional to P300 amplitudes. MFN amplitudes repeatedly indicated stronger conflicts during deceptive compared to honest responses for CITs as well as for a paradigm involving deception about attitudes. In CITs, P300 amplitudes indicated a greater salience of probe items, for which participants denied recognizing them, compared to irrelevant items requiring the honest response that they do know them. For deception about attitudes, P300 amplitudes indicated a stronger mental workload for items requiring a deceptive vs. honest response. The different patterns of P300 and MFN amplitudes for deceptive and honest response do not seem to rely on a forensic context of the deception paradigms. Unlike many previous deception studies, at least one condition of all studies of the thesis did not involve a crime and differing MFN and P300 amplitudes between deceptive and honest responses were also found in these non-forensic situations. Furthermore, study I involved a scenario in which participants were not exactly instructed what to do: They witnessed another person demonstrating a behavior causing a problem or demonstrated the same behavior on their own accord. The results of the study revealed that the P300- as well as MFN-based CIT effect was generalizable to a situation in which participants got to know the probe item during an uninstructed behavior.

Yet, for women, conflicts during deception, as indicated by MFN amplitudes, seem to be sensitive to the moral involvement of the person (witnessing vs. performing a behavior causing a problem). Accordingly, when relying on the MFN as a marker for deception, it has to be taken into account that an MFN-based CIT effect possibly does not occur for witnesses. Furthermore, it should be considered that in some situations, Machiavellianism can moderate the MFN-based CIT effect, in that conflicts during deception are less pronounced for individuals higher in Machiavellianism. In our study, this was the case for women witnessing another person performing a problematic behavior. In

contrast, the MFN-based CIT effect proved to occur for a morally negative (committing theft) as well as a morally positive situation (giving a present). Therefore, the MFN seems to be applicable as a marker for concealed knowledge for both morally negative and positive situations.

The patterns of P300 amplitudes in the deception paradigms were neither sensitive to moral involvement, nor moral valence, nor individual differences in Machiavellianism. The results suggest that the P300-based CIT effect holds true for multiple situations: morally positive and morally negative situations, witnesses as well as people performing a behavior causing a problem. Furthermore, the differing P300 amplitudes for probe vs. irrelevant items in CITs as well as for deceptive vs. honest responses about attitudes seem to be stable for individual differences in Machiavellianism. Overall, both P300- and MFN-based deception paradigms proved to be promising for detecting deception. Yet, it should be noted that MFN amplitudes can also be sensitive to the moral involvement of the examinee and that different patterns of P300 amplitudes occur in CITs and non-recognition deception tasks.

References

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A concealed information test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258–267. https://doi.org/10.1016/j.ijpsycho.2009.12.007

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170–177. https://doi.org/10.1016/j.tics.2004.02.010

Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: correlations with Machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality*, *14*(4), 359–368. https://doi.org/10.1002/1099-0984(200007/08)14:4<359::AID-PER382>3.0.CO;2-Y

Azizli, N., Atkinson, B. E., Baughman, H. M., Chin, K., Vernon, P. A., Harris, E., & Veselka, L. (2016). Lies and crimes: dark triad, misconduct, and high-stakes deception. *Personality and Individual Differences*, *89*, 34–39. https://doi.org/10.1016/j.paid.2015.09.034

Backbier, E. H. F., Hoogstraten, J., & Meerum Terwogt-Kouwenhoven, K. (1997). Situational determinants of the acceptability of telling lies. *Journal of Applied Social Psychology*, *27*(12), 1048–1062. https://doi.org/10.1111/j.1559-1816.1997.tb00286.x

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. D. (2010). Working memory. *Current Biology*, *20*(4), 136–140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. D., & Logie, R. H. (2016). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Pearson Education.

Bandura, A., Barbaranelli, C., Caprara, G. v, & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364–374. https://doi.org/10.1037/0022-3514.71.2.364

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: beyond errors and response conflict. *Psychophysiology, 42*(1), 33–42. https://doi.org/10.1111/j.1469-8986.2005.00258.x

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? *Advances in Experimental Social Psychology, 20*, 65–122. https://doi.org/10.1016/S0065-2601(08)60412-8

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In Theodore Millon & Melvin J. Lerner (Eds.), *Handbook of social psychology* (Vol. 5, pp. 463–506). John Wiley & Sons. https://doi.org/10.1002/0471264385.wei0519

Beauducel, A., Brocke, B., & Leue, A. (2006). Energetical bases of extraversion: effort, arousal, EEG, and performance. *International Journal of Psychophysiology*, *62*(2), 212–223. https://doi.org/10.1016/j.ijpsycho.2005.12.001

Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: an overview. *Frontiers in Psychology*, *3*, 342. https://doi.org/10.3389/fpsyg.2012.00342

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a meta-analytic review. *Journal of Applied Psychology, 88*(1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131

Bentham, J. (1789/1961). *An Introduction to the principles of morals and legislation*. Clarendon Press.

Bereczkei, T., Birkas, B., & Kerekes, Z. (2010). The presence of others, prosocial traits, Machiavellianism: a personality × situation approach. *Social Psychology, 41*(4), 238–245. https://doi.org/10.1027/1864-9335/a000032

Birbaumer, N., & Schmidt, R. F. (2010). *Biologische Psychologie [Biological Psychology]*. Springer.

Boksem, M. A. S., & De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Social Neuroscience, 5*(1), 118–128. https://doi.org/10.1080/17470910903202666

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments characterizations of deception. *Personality and Social Psychology Review, 10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Bradley, M. T., MacLaren, V., & Carle, S. B. (1996). Deception and non-deception in guilty knowledge and guilty actions polygraph tests. *Journal of Applied Psychology, 81*(2), 153–160. https://doi.org/10.1037/0021-9010.81.2.153

Brigham, J. C. (1999). What is forensic psychology, anyway? *Law and Human Behavior, 23*(3), 273–298. https://doi.org/10.1023/A:1022304414537

Bryant, E. (2008). Real Lies, white lies and gray lies: Towards a typology of deception. *Kaleidoscope: A Graduate Journal of Qualitative Communication Research, 7*, 23–48.

Camden, C., Motley, M. T., & Wilson, A. (1984). White lies in interpersonal communication: a taxonomy and preliminary investigation of social motivations. *Western Journal of Speech Communication, 48*(4), 309–325. https://doi.org/10.1080/10570318409374167

Campanella, S., Vanhoolandt, M. E., & Philippot, P. (2004). Emotional deficit in subjects with psychopathic tendencies as assessed by the Minnesota Multiphasic Personality Inventory-2: an event-related potentials study. *Neuroscience Letters, 373*(1), 26–31. https://doi.org/10.1016/j.neulet.2004.09.061

Carrión, R. E., Keenan, J. P., & Sebanz, N. (2010). A truth that's told with bad intent: an ERP study of deception. *Cognition, 114*(1), 105–110. https://doi.org/10.1016/j.cognition.2009.05.014

Carson, J. V., Lafree, G., & Dugan, L. (2012). Terrorist and non-terrorist criminal attacks by radical environmental and animal rights groups in the United States, 1970-2007. *Terrorism and Political Violence, 24*(2), 295–319. https://doi.org/10.1080/09546553.2011.639416

Carson, T. L. (2010). *Lying and deception: Theory and practice*. Oxford University Press.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. Academic Press. https://doi.org/10.2307/1958540

Christ, S. E., Essen, D. C. van, Watson, J. M., Brubaker, L. E., & Mcdermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex*, *19*(7), 1557–1566. https://doi.org/10.1093/cercor/bhn189

Coles, M. G. H., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. https://doi.org/10.1016/S0301-0511(01)00076-X

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235. https://doi.org/10.1037/a0031021

Cross, S. E., & Madson, L. (1997). Models of the self: self-construal theory and gender. *Psychological Bulletin*, *22*(1), 665–689. https://doi.org/10.1037/0033-2909.122.1.5

D'Agata, M. T. (2014). White lies. In T. R. Levine (Ed.), *Encyclopedia of Deception* (pp. 936–937). SAGE Publications, Inc. https://doi.org/https://www.doi.org/10.4135/9781483306902.n360

D'Aoust, A. M. (2013). In the Name of Love: marriage migration, governmentality, and technologies of love. *International Political Sociology*, *7*(3), 258–274. https://doi.org/10.1111/ips.12022

Decety, J., Chen, C., Harenski, C. L., & Kiehl, K. A. (2015). Socioemotional processing of morally-laden behavior and their consequences on others in forensic psychopaths. *Human Brain Mapping*, *36*(6), 2015–2026. https://doi.org/10.1002/hbm.22752

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trail EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology*, *71*(4), 703–716. https://doi.org/10.1037/0022-3514.71.4.703

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63–79. https://doi.org/10.1037/0022-3514.74.1.63

DePaulo, B. M., Kashy, D. A., Wyer, M. M., Kirkendol, S. E., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., & Charlton, K. (2003). Cues to Deception. *Psychological Bulletin 129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Dickter, C. L., & Bartholow, B. D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*(3), 189–198. https://doi.org/10.1093/scan/nsm012

Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: a systematic review and meta-analysis. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087347

Donchin, E., & Coles, M. G. H. (1978). Multivariate analysis of event-related potential data: a tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 555–572). U.S. Government printing office.

Dong, G., Wu, H., & Lu, Q. (2010). Attempting to hide our real thoughts: electrophysiological evidence from truthful and deceptive responses during evaluation. *Neuroscience Letters*, *479*(1), 1–5. https://doi.org/10.1016/j.neulet.2010.05.014

Doris, J., Stich, S., Philips, J., & Walmsley, L. (2018). Moral psychology: empirical approaches. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/moral-psych-emp

Driver, J. (2014). The history of utilitarianism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg, W. Damon, & M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 646–718). John Wiley & Sons.

Ekman, P. (2001). *Telling lies: clues to deceit in the marketplace, politics, and marriage.* W. W. Norton & Company.

Ekman, P. (2003). Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, *1000*, 205–221. https://doi.org/10.1196/annals.1280.010

Ekman, P. (2009). Lie catching and microexpressions. In C. Martin (Ed.), *The philosophy of deception* (pp. 118–139). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195327939.001.0001

Ekman, P., & Friesen, W. v. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*(1), 88–106. https://doi.org/10.1080/00332747.1969.11023575

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: a review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

Ennis, E., Vrij, A., & Chance, C. (2008). Individual differences and lying in everyday life. *Journal of Social and Personal Relationships*, *25*(1), 105–118. https://doi.org/10.1177/0265407507086808

Erat, S., & Gneezy, U. (2012). White Lies. *Management Science*, *58*(4), 723–733. https://doi.org/10.1287/mnsc.1110.1449

Fabiani, M., Gratton, G., & Federmeier, K. D. (2017). Event-related brain potentials: methods, theory, and applications. In *Handbook of psychophysiology* (Vol. 3, pp. 85–119). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.004

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology*, *23*(3), 298–308. https://doi.org/10.1111/j.1469-8986.1986.tb00636.x

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of cross-modal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Farwell, L. A. (2012). Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, *6*(2), 115–154. https://doi.org/10.1007/s11571-012-9192-2

Farwell, L. A., Richardson, D. C., & Richardson, G. M. (2013). Brain fingerprinting field studies comparing P300-MERMER and P300 brainwave responses in the detection of concealed information. *Cognitive Neurodynamics, 7*(4), 263–299. https://doi.org/10.1007/s11571-012-9230-0

Fehr, B., Samsom, D., & Paulhus, D. L. (1992). The construct of Machiavellianism: twenty years later. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Issue 9, pp. 77–116).

Foldvary-Schaefer, N., & Grigg-Damberger, M. M. (2012). Identifying interictal and ictal epileptic activity in polysomnograms. In M. M. Grigg-Damberger & N. R. Foldvary-Schaefer (Eds.), *Sleep-related epilepsy and Electroencephalography* (pp. 39–59). Elsevier.

Ford, E. B. (2006). Lie detection: historical, neuropsychiatric and legal dimensions. *International Journal of Law and Psychiatry*, *29*(3), 159–177. https://doi.org/10.1016/j.ijlp.2005.07.001

Forgas, J. P. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Attitudes and Social Cognition, 74*(3), 565–577.

Fullam, R. S., McKie, S., & Dolan, M. C. (2009). Psychopathic traits and deception: functional magnetic resonance imaging study. *British Journal of Psychiatry, 194*(3), 229–235. https://doi.org/10.1192/bjp.bp.108.053199

Gamer, M. (2010). Does the guilty actions test allow for differentiating guilty participants from informed innocents? A re-examination. *International Journal of Psychophysiology, 76*(1), 19–24. https://doi.org/10.1016/j.ijpsycho.2010.01.009

Gamer, M., & Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology, 47*(2), 355–364. https://doi.org/10.1111/j.1469-8986.2009.00933.x

Gamer, M., & Berti, S. (2012). P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses. *Frontiers in Human Neuroscience, 6*, 308. https://doi.org/10.3389/fnhum.2012.00308

Ganis, G., & Keenan, J. P. (2009). The cognitive neuroscience of deception. *Social Neuroscience, 4*(6), 465–472. https://doi.org/10.1080/17470910802507660

Gao, Y., & Raine, A. (2009). P3 event-related potential impairments in antisocial and psychopathic individuals: a meta-analysis. *Biological Psychology*, *82*(3), 199–210. https://doi.org/10.1016/j.biopsycho.2009.06.006

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., & Knight, R. T. (2000). Prefrontal – cingulate interactions in action monitoring. *Nature Neuroscience, 3*(5), 516-520. https://doi.org/10.1038/74899

Gehring, W. J., & Willoughby, A. R. (2004). Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain: Current opinions on performance monitoring* (pp. 14–20). Max Planck Institute for Human Cognitive and Brain Sciences.

Geis, F. L., & Moon, T. H. (1981). Machiavellianism and deception. *Journal of Personality and Social Psychology*, *41*(4), 766–775. https://doi.org/10.1037/0022-3514.41.4.766

Geis, F. L., Weinheimer, S., & Berger, D. (1970). Playing legislature: Cool heads and hot issues. In *Studies in Machiavellianism* (pp. 190–209). Academic Press.

Gershoff, A. D., Mukherjee, A., & Mukhopadhyay, A. (2008). What's not to like? Preference asymmetry in the false consensus effect. *Journal of Consumer Research, 35*(1), 119–125. https://doi.org/10.1086/524416

Ghosh, D., & Crain, T. L. (1995). Ethical standards, attitudes toward risk, and intentional noncompliance: an experimental investigation. *Journal of Business Ethics, 14*(5), 353–365. https://doi.org/10.1007/BF00872098

Gibbons, H., Schnuerch, R., Wittinghofer, C., Armbrecht, A. S., & Stahl, J. (2018). Detection of deception: Event-related potential markers of attention and cognitive control during intentional false responses. *Psychophysiology*, *55*(6), 1–22. https://doi.org/10.1111/psyp.13047

Gilligan, C. (1977). In a different voice: Women's conceptions of self and of morality. *Harvard Educational Review*, *47*(4), 481–517. https://doi.org/10.17763/haer.47.4.g6167429416hg5l0

Gilligan, C., & Attanucci, J. (1988). Two moral orientations: gender differences and similarities. *Merrill-Palmer Quarterly*, *34*(3), 223–237.

Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs*, *132*(3), 197–214. https://doi.org/10.3200/MONO.132.3.197-214

Gozna, L. F., Vrij, A., & Bull, R. (2001). The impact of individual differences on perceptions of lying in everyday life and in a high stake situation. *Personality and Individual Differences*, *31*(7), 1203–1216. https://doi.org/10.1016/S0191-8869(00)00219-1

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1097/00005053-199107000-00016

Granhag, P. A., & Strömwall, L. A. (2004). Research on deception detection: past and present. *The Detection of Deception in Forensic Contexts*, 3–12. https://doi.org/10.1017/CBO9780511490071.001

Granhag, P. A., & Vrij, A. (2005). Deception detection. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: an empirical perspective* (pp. 43–93). The Guilford Press.

Granhag, P. A., Vrij, A., & Verschuere, B. (2015). *Detecting deception: current challenges and cognitive approaches*. Wiley.

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: a fundamental template unifying moral judgment. *Psychological Inquiry*, *23*(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Gray, K., & Wegner, D. M. (2011). Dimensions of moral emotions. *Emotion Review*, *3*(3), 258–260. https://doi.org/10.1177/1754073911402388

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*(1), 65–72. https://doi.org/10.1111/j.1745-6916.2008.00063.x

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221. https://doi.org/10.1111/j.1559-1816.2001.tb02489.x

Haidt, J., & Kesebir, S. (2008). Morality. In *Handbook of Social Psychology* (pp. 797–832). https://doi.org/10.5840/traddisc2009/201036345

Hancock, J. T., Thom-Santelli, J., & Ritchie, T. (2004). Deception and design: The impact of communication technology on lying behavior. *Proceedings of the SIGCHI Conference on Human Factors in Computing*, *6*, 129–134. https://doi.org/10.1145/985692.985709

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: I Studies in deceit*. Macmillan.

Hayashi, A., Abe, N., Fujii, T., Ito, A., Ueno, A., Koseki, Y., Mugikura, S., Takahashi, S., & Mori, E. (2014). Dissociable neural systems for moral judgment of anti- and pro-social lying. *Brain Research*, *1556*, 46–56. https://doi.org/10.1016/j.brainres.2014.02.011

Hegarty, W. H., & Sims, H. P. (1978). Some determinants of unethical decision behavior: an experiment. *Journal of Applied Psychology*, *63*(4), 451–457. https://doi.org/10.1037/0021-9010.63.4.451

Hegarty, W. H., & Sims, H. P. (1979). Organizational philosophy, policies, and objectives related to unethical decision behavior: a laboratory experiment. *Journal of Applied Psychology*, *64*(3), 331–338. https://doi.org/10.1037//0021-9010.64.3.331

Henning, H. J., & Six, B. (1977). Konstruktion einer Machiavellismus-Skala [Construction of a Machiavellianism Scale]. *Zeitschrift Für Sozialpsychologie*, *8*, 185–198.

Hruby, T., & Marsalek, P. (2003). Event-related potentials- the P3 wave. *Acta Neurobiologiae Experimentalis*, *63*, 55–63. https://doi.org/12784933

Hu, X., & Mai, X. (2021). Social value orientation modulates fairness processing during social decision-making: Evidence from behavior and brain potentials. *Social Cognitive and Affective Neuroscience*, *12*(11), 1730–1739. https://doi.org/10.1093/scan/nsab032

Hu, X., & Rosenfeld, J. P. (2012). Combining the P300-complex trial-based concealed information test and the reaction time-based autobiographical implicit association test in concealed memory detection. *Psychophysiology*, *49*(8), 1090–1100. https://doi.org/10.1111/j.1469-8986.2012.01389.x

Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self- and other-referential information: an ERP study. *Brain Research, 1369*, 149–157. https://doi.org/10.1016/j.brainres.2010.10.106

Iacono, W. G. (2014). Forensic application of event-related brain potentials to detect guilty knowledge. In P. A. Granhag, B. Verschuere, & A. Vrij (Eds.), *Detecting Deception* (pp. 81–105).

Jaffe, S., & Hyde, J. S. (2000). Gender differences in moral orientation. *Psychological Bulletin*, *126*(5), 703–726. https://doi.org/10.1037/0033-2909.126.5.703

Jang, K.-W., Kim, D.-Y., Cho, S., & Lee, J.-H. (2013). Effects of the combination of P3-based GKT and reality monitoring on deceptive classification. *Frontiers in Human Neuroscience*, *7*, 18. https://doi.org/10.3389/fnhum.2013.00018

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*(4), 367–384. https://doi.org/doi.org/10.1111/j.1469-8986.1986.tb00649.x

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: review and synthesis. *Advances in Psychophysiology*, *3*, 69–137.

Johnson, R. (2014). The neural basis of deception and credibility assessment: a cognitive neuroscience perspective. In *Credibility Assessment: Scientific Research and Applications* (pp. 217–300). Elsevier. https://doi.org/10.1016/B978-0-12-394433-7.00006-3

Johnson, R., Barnhardt, J., & Zhu, J. (2003). The deceptive response: Effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology*, *64*(3), 217–253. https://doi.org/10.1016/j.biopsycho.2003.07.006

Johnson, R., Barnhardt, J., & Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia*, *42*(7), 878–901. https://doi.org/10.1016/j.neuropsychologia.2003.12.005

Johnson, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Cognitive Brain Research*, *24*(3), 386–404. https://doi.org/10.1016/j.cogbrainres.2005.02.011

Johnson, R., Henkell, H., Simon, E., & Zhu, J. (2008). The self in conflict: The role of executive processes during truthful and deceptive responses about attitudes. *NeuroImage*, *39*(1), 469–482. https://doi.org/10.1016/j.neuroimage.2007.08.032

Jonason, P. K., Lyons, M., Baughman, H. M., & Vernon, P. A. (2014). What a tangled web we weave: the dark triad traits and deception. *Personality and Individual Differences*, *70*, 117–119. https://doi.org/10.1016/j.paid.2014.06.038

Jones, D. N., & Mueller, S. M. (2021). Is Machiavellianism dead or dormant? The perils of researching a secretive construct. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-020-04708-w

Jones, D. N., & Paulhus, D. L. (2009). Machavellianism. In M. R. Levy & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–109). The Guilford Press.

Jones, D. N., & Paulhus, D. L. (2014). Introducing the short dark triad (SD3): a brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Jutai, J. W., Hare, R. D., & Connolly, J. F. (1987). Psychopathy and event-related brain Potentials (ERPs) associated with attention to speech stimuli. *Personality and Individual Differences*, *8*(2), 175–184. https://doi.org/10.1016/0191-8869(87)90172-3

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten [Groundwork of the metaphysics of morals]*. Anakonda. https://korpora.zim.uni-duisburg-essen.de/Kant/aa04/421.html

Kant, I. (1797). Über ein vermeintliches Recht aus Menschenliebe zu lügen [On a supposed right to lie from philanthropy]. In *Metaphysik der Sitten* (pp. 637–643). Suhrkamp. https://korpora.zim.uni-duisburg-essen.de/Kant/aa08/425.html

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*, *47*(5), 888–904. https://doi.org/10.1111/j.1469-8986.2010.01009.x.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology, 70*(5), 1037–1051. https://doi.org/10.1037/0022-3514.70.5.1037

Kenward, B., & Dahl, M. (2011). Preschoolers distribute scarce resources according to the moral valence of recipients' previous actions. *Developmental Psychology, 47*(4), 1054–1064. https://doi.org/10.1037/a0023869

Kiehl, K. A., Bates, A. T., Laurens, K. R., Hare, R. D., & Liddle, P. F. (2006). Brain potentials implicate temporal lobe abnormalities in criminal psychopaths. *Journal of Abnormal Psychology, 115*(3), 443–453. https://doi.org/10.1037/0021-843X.115.3.443

Kiehl, K. A., Hare, R. D., Liddle, P. F., & McDonald, J. J. (1999). Reduced P300 responses in criminal psychopaths during a visual oddball task. *Biological Psychiatry, 45*(11), 1498–1507. https://doi.org/10.1016/S0006-3223(98)00193-0

Killen, M., & Smetana, J. G (2014). *Handbook of moral development*. Psychology Press.

Kireev, M. v, Pakhomov, S. v, & Medvedev, S. v. (2008). Cerebral mechanisms of error detection during deceptive responses in the normal state and under the influence of alcohol. *Human Physiology, 34*(2), 141–149. https://doi.org/10.1134/S0362119708020023

Kleinmuntz, B., & Szucko, J. J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist, 39*(7), 766–776. https://doi.org/10.1037/0003-066X.39.7.766

klein Selle, N., Gueta, C., Harpaz, Y., Deouell, L. Y., & Ben-Shakhar, G. (2021). Brain-based concealed memory detection is driven mainly by orientation to salient items. *Cortex, 136*, 41–55. https://doi.org/10.1016/j.cortex.2020.12.010

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E. H., & Ben-Shakhar, G. (2016). Orienting versus inhibition in the concealed information test: Different cognitive processes drive different physiological measures. *Psychophysiology, 53*(4), 579–590. https://doi.org/10.1111/psyp.12583

Kohlberg, L. (2008). The development of children's orientations toward a moral order. *Human Development, 51*(1), 8–20. https://doi.org/10.1159/000112530

Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth' In P. A. Granhag & L. A. Strömwall (Eds.), *The Detection of Deception in Forensic*

Contexts (pp. 41–63). Cambridge University Press. https://doi.org/10.1017/CBO9780511490071.003

Lamb, M. E., I., H., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Wiley.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping*, *26*(4), 262–272. https://doi.org/10.1002/hbm.20191

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., O'brien, C. P., & Childress, A. R. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *NeuroImage*, *15*(3), 727–732. https://doi.org/10.1006/nimg.2001.1003

Lavoie, J., Leduc, K., Crossman, A. M., & Talwar, V. (2016). Do as I say and not as I think: parent socialisation of lie-telling behaviour. *Children & Society*, *30*(4), 253–264. https://doi.org/10.1111/chso.12139

Lee, K., & Ashton, M. C. (2005). Psychopathy, Machiavellianism, and narcissism in the Five-Factor Model and the HEXACO model of personality structure. *Personality and Individual Differences*, *38*(7), 1571–1582. https://doi.org/10.1016/j.paid.2004.09.016

Lee, K., & Ross, H. J. (1997). The concept of lying in adolescents and young adults: Testing Sweetser's folkloristic model. *Merrill-Palmer Quarterly*, *43*(2), 255–270.

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, *44*, 894–904. https://doi.org/10.1111/j.1469-8986.2007.00566.x

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2009). Use of event-related brain potentials (ERPs) to assess eyewitness accuracy and deception. *International Journal of Psychophysiology*, *73*(3), 218–225. https://doi.org/10.1016/j.ijpsycho.2009.03.003

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: theory of mind and moral judgment. *Psychological Science*, *17*(5), 421–427. https://doi.org/10.1111/j.1467-9280.2006.01722.x

Leue, A., & Beauducel, A. (2015). Effects of injustice sensitivity and sex on the P3 amplitude during deception. *Biological Psychology, 109,* 29–36. https://doi.org/10.1016/j.biopsycho.2015.04.004

Leue, A., & Beauducel, A. (2019). A meta-analysis of the P3 amplitude in tasks requiring deception in legal and social contexts. *Brain and Cognition, 135,* 103564. https://doi.org/10.1016/j.bandc.2019.05.002

Leue, A., Lange, S., & Beauducel, A. (2012). "Have you ever seen this face?" - Individual differences and event-related potentials during deception. *Frontiers in Psychology, 3,* 570. https://doi.org/10.3389/fpsyg.2012.00570

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53,* 107–117. https://doi.org/10.1016/j.jesp.2014.03.005

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes, 126,* 88–106. https://doi.org/10.1016/j.obhdp.2014.10.007

Lindskold, S., & Han, G. (1986). Intent and the judgment of lies. *Journal of Social Psychology*, *126*(1), 129–130. https://doi.org/10.1080/00224545.1986.9713581

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 17–32). MIT Press. https://doi.org/10.1371/journal.pcbi.0020012

Luck, S. J. (2014). *An introduction to the event-related potential technique*. The MIT Press.

Lui, M. A., & Rosenfeld, J. P. (2008). Detection of deception about multiple, concealed, mock crime items, based on a spatial-temporal analysis of ERP amplitude and scalp distribution. *Psychophysiology*, *45*(5), 721–730. https://doi.org/10.1111/j.1469-8986.2008.00683.x

Lu, Y., Rosenfeld, J. P., Deng, X., Zhang, E., Zheng, H., Yan, G., Ouyang, D., & Hayat, S. Z. (2018). Inferior detection of information from collaborative versus individual crimes based on a P300 concealed information test. *Psychophysiology*, *55*(4), 1–13. https://doi.org/10.1111/psyp.13021

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. https://doi.org/10.1037/h0046060

Machiavelli, N. (2009). *Der Fürst*. Nikol.

Mackey, D. A., & Elvey, K. M. (2021). *Society, Ethics and the Law: A Reader*. Jones & Bartlett Learning.

McLeod, B. A., & Genereux, R. L. (2008). Predicting the acceptability and likelihood of lying: the interaction of personality with type of lie. *Personality and Individual Differences*, *45*(7), 591–596. https://doi.org/10.1016/j.paid.2008.06.015

Meek, S. W., Phillips, M. C., Boswell, C. P., & Vendemia, J. M. C. (2013). Deception and the misinformation effect: an event-related potential study. *International Journal of Psychophysiology*, *87*(1), 81–87. https://doi.org/10.1016/j.ijpsycho.2012.11.004

Meijer, E. H., klein Selle, K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the concealed information test: a meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51(9)*, 879–904. https://doi.org/10.1111/psyp.12239

Meijer, E. H., Smulders, F. T. Y., Merckelbach, H. L. G. J., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology*, *66*(3), 231–237. https://doi.org/10.1016/j.ijpsycho.2007.08.001

Meijer, E. H., & Verschuere, B. (2015). The polygraph: current practices and new approaches. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception* (pp. 59–81). Wiley. https://doi.org/10.1002/9781118510001

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, *53*(5), 593–604. https://doi.org/10.1111/psyp.12609

Meixner, J. B. (2018). Admissibility and consitutional issues of the concealed information test in American courts: an update. In J. P. Rosenfeld (Ed.), *Detecting concealed information and deception* (pp. 405–431). Academic Press.

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology*, *48*(2), 149–154. https://doi.org/10.1111/j.1469-8986.2010.01050.x

Miller, A. R., & Rosenfeld, J. P. (2004). Response-specific scalp distributions in deception detection and ERP correlates of psychopathic personality traits. *Journal of Psychophysiology*, *18*(1), 13–26. https://doi.org/10.1027/0269-8803.18.1.13

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, *80*(1), 75–78. https://doi.org/10.1016/j.ijpsycho.2011.02.003

Mill, J. S. (1863/2001). *Utilitarianism*. Batoche Books.

Milnik, V. (2012). *Elektrophysiologie in der Praxis [Electroencephalography in practice]*. Elsevier.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *100*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moore, C., Detert, J. R., Treviño, L. K., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: moral disengagement and unethical organisational behaviour. *Personnel Psychology*, *65*(1), 1–48. https://doi.org/10.1111/j.1744-6570.2011.01237.x

Mudrack, P. E. (1993). An investigation into the acceptability of workplace behaviors of a dubious ethical nature. *Journal of Business Ethics*, *12*(7), 517–524. https://doi.org/10.1007/BF00872373

Mudrack, P. E., & Mason, E. S. (1995). More on the acceptability of workplace behaviors of a dubious ethical nature. *Psychological Reports*, *76*, 639–648. https://doi.org/10.1007/BF00872373

Murphy, P. R. (2012). Attitude, Machiavellianism and the rationalization of misreporting. *Accounting, Organizations and Society*, *37*(4), 242–259. https://doi.org/10.1016/j.aos.2012.04.002

Nahari, T., Breska, A., Elber, L., klein Selle, N., & Ben-Shakhar, G. (2017). The external validity of the concealed information test: the effect of choosing to commit a mock crime. *Applied Cognitive Psychology*, *31*(1), 81–90. https://doi.org/10.1002/acp.3304

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience and Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003

Olson, J. M., Rosenfeld, J. P., & Perrault, E. (2019). Deleterious effects of probe-related versus irrelevant targets on the "CIT effect" in the P300- and RT-based three-stimulus protocol for detection of concealed information. *Psychophysiology, 56*(12), 1–10. https://doi.org/10.1111/psyp.13459

Palmer, B., Nasman, V. T., & Wilson, G. F. (1994). Task decision difficulty: effects on ERPs in a same-different letter classification task. *Biological Psychology*, *38*, 199–214. https://doi.org/10.1016/0301-0511(94)90039-6

Panasiti, M. S., Pavone, E. F., Mancini, A., Merla, A., Grisoni, L., & Aglioti, S. M. (2014). The motor cost of telling lies: electrocortical signatures and personality foundations of spontaneous deception. *Social Neuroscience*, *9*(6), 573–589. https://doi.org/10.1080/17470919.2014.934394

Pape, H. C. (2009). Wachheit und Schlaf: Rhythmen des Gehirns im Muster des Elektroenzephalogramms [Awakeness and sleep: Rhythms of the brain in the electroencephalogramm]. In R. Klinke, H. C. Pape, A. Kurtz, & S. Silbernagel (Eds.), *Physiologie* (pp. 850–862). Georg Thieme Verlag.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, *23*(10), 3162–3180. https://doi.org/10.1162/jocn_a_00017

Peng, W. (2019). EEG Preprocessing and denoising. In L. Hu & Z. Zhang (Eds.), *EEG signal processing and feature extraction* (pp. 71–89). Springer. https://doi.org/10/1.007/978-13-9113-2

Pesciarelli, F., Leo, I., & Serafini, L. (2021). Electrophysiological correlates of unconscious processes of race. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-91133-2

Peterson, C. C. (1996). Deception in intimate relationships. *International Journal of Psychology*, *31*(6), 279–288.

Peterson, C. C., Peterson, J. L., & Seeto, D. (1983). Developmental changes in ideas about lying. *Child Development*, *54*(6), 1529–1535.

Pfister, R., Foerster, A., & Kunde, W. (2014). Pants on fire: The electrophysiological signature of telling a lie. *Social Neuroscience*, *9*(6), 562–572. https://doi.org/10.1080/17470919.2014.934392

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, *41*(95), 103–146. https://doi.org/10.1016/0301-0511(95)05130-9

Posner, M. I., & Deheane, S. (1994). Attentional Networks. *Trends in Neurosciences*, *17*(2), 75–79. https://doi.org/10.1016/B978-008045046-9.00204-7

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*(3), 506–540. https://doi.org/10.1037/0033-2909.89.3.506

Rachels, J. (2013). Naturalism. In H. LaFolette & I. Persson (Eds.), *The blackwell guide to ethical theory* (pp. 74–91). Blackwell.

Raine, A., & Venables, P. H. (1988). Enhanced P3 evoked potentials and longer recovery times in psychopaths. *Psychophysiology*, *25*(2), 30–38.

Robinson, W. P. (1994). Reactions to falsifications in public and interpersonal contexts. *Journal of Language and Social Psychology*, *13*(4), 497–513.

Rokeach, M. (1973). *The nature of human values*. Free Press.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the concealed information test* (pp. 63–89). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196

Rosenfeld, J. P. (2020). P300 in detecting concealed information and deception: a review. *Psychophysiology*, *57*(7), 1–12. https://doi.org/10.1111/psyp.13362

Rosenfeld, J. P., Angell, A., Johnson, M. M., & Qian, J.-H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *28*(3), 319–335. https://doi.org/https://doi.org/10.1111/j.1469-8986.1991.tb02202.x

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology*, *60*(3), 251–259. https://doi.org/10.1016/j.ijpsycho.2005.06.002

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J. B., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*(2), 118–134. https://doi.org/10.1016/j.ijpsycho.2013.08.012

Rosenfeld, J. P., Hu, X., & Pederson, K. (2012). Deception awareness improves P300-based deception detection in concealed information tests. *International Journal of Psychophysiology*, *86*(1), 114–121. https://doi.org/10.1016/j.ijpsycho.2012.06.007

Rosenfeld, J. P., Labkovsky, E., Davydova, E., Ward, A. C., & Rosenfeld, L. (2017). Financial incentive does not affect P300 (in response to certain episodic and semantic probe stimuli) in the complex trial protocol (CTP) version of the concealed information test (CIT) in detection of malingering. *Psychophysiology*, *54*(5), 764–772. https://doi.org/10.1111/psyp.12835

Rosenfeld, J. P., Ozsan, I., & Ward, A. C. (2017). P300 amplitude at Pz and N200/N300 latency at F3 differ between participants simulating suspect versus witness roles in a mock crime. *Psychophysiology*, *54*(4), 640–648. https://doi.org/10.1111/psyp.12823

Rosenfeld, J. P., Sitar, E., Wasserman, J., & Ward, A. C. (2018). Moderate financial incentive does not appear to influence the P300 concealed information test (CIT) effect in the complex trial protocol (CTP) version of the CIT in a forensic scenario, while affecting P300 peak latencies and behavior. *International Journal of Psychophysiology*, *125*, 42–49. https://doi.org/https://doi.org/10.1016/j.ijpsycho.2018.02.006

Rosenfeld, J. P., Ward, A. C., Thai, M., & Labkovsky, E. (2015). Superiority of pictorial versus verbal presentation and initial exposure in the P300-based complex trial protocol for concealed memory detection. *Applied Psychophysiology Biofeedback*, *40*(2), 61–73. https://doi.org/10.1007/s10484-015-9275-z

Rothbart, M. K., Hanley, D., & Albert, M. (1986). Gender differences in moral reasoning. *Sex Roles*, *15*(11–12), 645–653. https://doi.org/10.1007/BF00288220

Sai, L., Li, H., Wang, C., Rosenfeld, J. P., Lin, X., & Fu, G. (2020). Feedback does not influence the recognition-related P300 in a novel concealed information test while feedback-evoked P300 shows promising diagnostic accuracy. *International Journal of Psychophysiology, 157*, 32–41. https://doi.org/10.1016/j.ijpsycho.2020.08.003

Sai, L., Lin, X., Hu, X., & Fu, G. (2014). Detecting concealed information using feedback related event-related brain potentials. *Brain and Cognition, 90*, 142–150. https://doi.org/10.1016/j.bandc.2014.06.012

Sakalaki, M., Richardson, C., & Thépaut, Y. (2007). Machiavellianism and economic opportunism. *Journal of Applied Social Psychology, 37*(6), 1181–1190. https://doi.org/10.1111/j.1559-1816.2007.00208.x

Sanei, S., & Chambers, J. (2007). *EEG signal processing*. John Wiley & Sons.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science*, *19*(8), 772–780. https://doi.org/10.1111/j.1467-9280.2008.02156.x

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Scheuble, V., & Beauducel, A. (2020a). Cognitive processes during deception about attitudes revisited: a replication study. *Social Cognitive and Affective Neuroscience*, *15*(8), 839–848. https://doi.org/10.1093/scan/nsaa107

Scheuble, V., & Beauducel, A. (2020b). Individual differences in ERPs during deception: Observing vs. demonstrating behavior leading to a small social conflict. *Biological Psychology, 150*, 107830. https://doi.org/10.1016/j.biopsycho.2019.107830

Scheuble, V., Mildenberger, M., & Beauducel, A. (2021). The P300 and MFN as indicators of concealed knowledge in situations with negative and positive moral valence. *Biological Psychology, 162*, 108093. https://doi.org/10.1016/j.biopsycho.2021.108093

Seifert, J. (2005). *Ereigniskorrelierte EEG-Aktivität [Event-related EEG activity]*. Pabst Science Publisher.

Seiter, J. S., & Bruschke, J. (2007). Deception and emotion: The effects of motivation, relationship type, and sex on expected feelings of guilt and shame following acts of deception in United States and Chinese samples. *Communication Studies*, *58*(1), 1–16. https://doi.org/10.1080/10510970601168624

Seiter, J. S., Bruschke, J., & Bai, C. (2009). The acceptability of deception as a function of perceivers' culture, deceiver's intention, and deceiver-deceived relationship. *Western Journal of Communication*, *66*(2), 158–180. https://doi.org/10.1080/10570310209374731

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: three studies of self-reported lies. *Human Communication Research*, *36*(1), 2–25. https://doi.org/10.1111/j.1468-2958.2009.01366.x

Shenhav, A., Cohen, J., & Botvinick, M. M. (2016). Dorsal anterior cingulate cortex and the value of control. *Nature Neuroscience*, *19*(10), 1280–1285. https://doi.org/10.1038/nn.4382

Sims, R. L. (2002). Support for the use of deception within the work environment: a comparison of Israeli and United States employee attitudes. *Journal of Business Ethics*, *35*(1), 27–34. https://doi.org/10.1023/A:1012755801190

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in Cognitive Sciences*, *12*(2), 48–53. https://doi.org/10.1016/j.tics.2007.11.008

Smith, L. (2007). *ActiveTwo System*. https://brain.mcmaster.ca/file/ActiveTwo.Rev.2007.pdf

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms. *International Journal of Psychophysiology*, *40*(2), 173–180. https://doi.org/10.1016/S0167-8760(00)00154-9

Speckmann, E.-J., Elger, C. E., & Gorji, A. (2012). Neurophysiologic basis of EEG and DC potentials. In D. L. Schomer & F. H. L. da Silva (Eds.), *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields* (6th ed., pp. 17–31). Wolters Kluwer.

St Augustine. (1995). On lying [De Mendacio]. In Philipp Schaff (Ed.), *Nicene and post-nicene fathers* (pp. 457–477). Hendrickson Publishers.

Suchotzki, K., Crombez, G., Smulders, F. T. Y., Meijer, E. H., & Verschuere, B. (2015). The cognitive mechanisms underlying deception: An event-related potential study. *International Journal of Psychophysiology, 95*(3), 395–405. https://doi.org/10.1016/j.ijpsycho.2015.01.010

Suchotzki, K., Verschuere, B., van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428–453. https://doi.org/10.1037/bul0000087

Sweetser, E. E. (1987). The definition of lie: An examination of the folk models underlying a semantic prototype. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 43–66). Cambridge University Press.

Tays, W. J., Dywan, J., Mathewson, K. J., & Segalowitz, S. J. (2008). Age differences in target detection and interference resolution in working memory: an event-related potential study. *Journal of Cognitive Neuroscience*, *20*(12), 2250–2262. https://doi.org/10.1162/jocn.2008.20158

Trovillo, P. v. (1939). A history of lie detection. *Journal of Criminal Law and Criminology (1931-1951)*, *29*(6), 848. https://doi.org/10.2307/1136489

Turiel, E. (1983). *The development of social knowledge: morality and convention*. Cambridge University Press.

Turner, R. E., Edgley, C., & Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology, 11*, 69–89.

Vendemia, J. M. C., & Buzan, R. F. (2005). Neural mechanisms of deception and response congruity in a visual two-stimulus paradigm with motor response. *Polygraph*, *34*(1), 24–46.

Verschuere, B., Rosenfeld, J. P., Winograd, M. R., Labkovsky, E., & Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal and Criminological Psychology*, *14*(2), 253–262. https://doi.org/10.1348/135532508X384184

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the “error negativity” specific to errors? *Biological Psychology*, *51*(2–3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Wacker, J. (2017). Increasing the reproducibility of science through close cooperation and forking path analysis. *Frontiers in Psychology*, *8*, 1332. https://doi.org/10.3389/fpsyg.2017.01332

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying to questions: Response time as a cue to deception. *Applied Cognitive Psychology*, *17*(7), 755–774. https://doi.org/10.1002/acp.914

Ward, A. C., Rosenfeld, J. P., Sitar, E. J., & Wasserman, J. D. (2020). The effect of retroactive memory interference on the P300-based Complex Trial Protocol (CTP). *International Journal of Psychophysiology*, *147*, 213–223. https://doi.org/10.1016/j.ijpsycho.2019.10.016

Wark, G. R., & Krebs, D. L. (1996). Gender and dilemma differences in real-life moral judgment. *Developmental Psychology*, *32*(2), 220–230. https://doi.org/10.1037/0012-1649.32.2.220

Weber, J. (1993). Exploring the relationship between personal values and moral reasoning. *Human Relations*, *46*(4), 435–463. https://doi.org/10.1177/001872679304600402

What is the function of the CMS and DRL electrodes. (n.d.). https://www.biosemi.com/faq/cms&drl.htm

Winograd, M. R., & Rosenfeld, J. P. (2011). Mock crime application of the complex trial protocol (CTP) P300-based concealed information test. *Psychophysiology*, *48*(2), 155–161. https://doi.org/10.1111/j.1469-8986.2010.01054.x

Winograd, M. R., & Rosenfeld, J. P. (2014). The impact of prior knowledge from participant instructions in a mock crime P300 concealed information test. *International Journal of Psychophysiology*, *94*(3), 473–481. https://doi.org/10.1016/j.ijpsycho.2014.08.002

Wissing, B. G., & Reinhard, M. A. (2019). The dark triad and deception perceptions. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01811

Wynia, M. K., Cummins, D. S., VanGeest, J. B., & Wilson, I. B. (2000). Physician manipulation of reimbursement rules for patients. *Jama*, *283*(14), 1858. https://doi.org/10.1001/jama.283.14.1858

Xu, F., Luo, Y. C., Fu, G., & Lee, K. (2009). Children's and adults' conceptualization and evaluation of lying and truth-telling. *Infant and Child Development*, *18*(6), 307–322. https://doi.org/10.1002/icd.631

Yamada, T., & Meng, E. (2018). *Practical Guide for Clinical Neurophysiologic Testing: EEG*. Lippincott Williams & Wilkins.

You, D., Maeda, Y., & Bebeau, M. J. (2011). Gender differences in moral sensitivity: a meta-analysis. *Ethics and Behavior*, *21*(4), 263–282. https://doi.org/10.1080/10508422.2011.585591

Zschocke, S. (2002). *Klinische Elektroenzephalographie [Clinical electroencephalography]*. Springer. https://doi.org/10.1007/978-3-662-08106-8

List of Abbreviations

CIT	Concealed information test
CMS	Common mode sense active electrode
DRL	Driven right leg passive electrode
ERP	Event-related potentials
fMRI	Functional magnetic resonance imaging
GAT	Guilty action test
IAT	Implicit association test
MFN	Medial frontal negativity
PET	Positron emissions topography
SVA	Statement validity assessment

List of Figures and Tables

Figures

Tables

Appendix

The present thesis is based on three publications. To avoid copyright violations, the articles have not been included in this thesis. They can be found online via the following references.

Article I: Scheuble, V. & Beauducel, A. (2020). Individual differences in ERPs during deception: Observing vs. demonstrating behavior leading to a small social conflict. *Biological Psychology*, 150. https://doi.org/10.1016/j.biopsycho.2019.107830

Article II: Scheuble, V. & Beauducel, A. (2020). Cognitive Processes during deception about attitudes revisited: a replication study. *Social Cognitive and Affective Neuroscience*, 1-10. https://doi.org/10.1093/scan/nsaa107

Article III: Scheuble, V.**,** Mildenberger, M., & Beauducel, A. (2021). The P300 and MFN as indicators of concealed knowledge in situations with negative and positive moral valence. *Biological Psychology*, 108093. https://doi.org/10.1016/j.biopsycho.2021.108093