BABEŞ-BOLYAI UNIVERSITY

FACULTY OF PSYCHOLOGY AND EDUCATIONAL SCIENCES

DOCTORAL SCHOOL "APPLIED COGNITIVE PSYCHOLOGY"

(Motivated) False Beliefs:

Individual and Contextual Factors Shaping Self-Deception and Memory Errors

SUMMARY OF THE PH.D. THESIS

Author: PhD candidate Mara Georgiana Moldoveanu

Scientific Supervisors: Prof. dr. habil. Laura Visu-Petra,

Babeș-Bolyai University, Romania

Prof. dr. Henry Otgaar

Maastricht University, the Netherlands

Cluj-Napoca

2025

TABLE OF CONTENTS

Keywords: self-deception; false beliefs; social desirability; self-enhancement**;** motivated false beliefs; autobiographical beliefs; event plausibility; event frequency; moral self-concept; memory erros; aggression; lie acceptability;

ACKNOWLEDGEMENTS

"Tell me, what is it you plan to do with your one wild and precious life?"Mary Oliver

First of all, I would love to say thank you and to express my gratitude to my supervisor, Laura. I also want to express my gratitude to my supervisor, Henry. Both Laura and Henry have profoundly shaped my academic, professional, and personal growth. Laura, by inspiring me with her empathy, kindness, and collaborative spirit, and Henry, by exemplifying integrity, support, and a balanced, caring approach to research and mentorship, for which I am deeply grateful. Thank you!

I want to thank my colleagues from Babeș-Boyai University and my Advisory Committee. Also, I am deeply grateful to Prof Dr Harald Merckelbach, our colleagues from the Forensic Psychology Department at Maastricht. my colleagues at KU Leuven. I also acknowledge my mum, my siblings (Multumesc Simona, Petre, Mihaela, Dorina), and my soul sister Ani. Also, thank you, Lavinia, Cesc Joana, Mari, Lien and Simon, Raluca, Guadalupe, my communities in Cluj-Napoca and Catalunya. Thank you! Also, this past year would not have been as inspiring and challenging if it weren't for my colleagues from the labs I have been part of. To my RIDDLE Lab colleagues: Thank you, Maria and Narcisa, Lili, Andreea, Simina, Daria and Lexi. To my CELL peeps: thank you, Charlotte and Paul, Tamara, Ivan and Fabiana, Yikang, Nurul, Diandra, Elsa, Caro, Ophelia, and Anne.I also want to thank the students I worked with along the way (especially Gabi, Diana, Eugen, and Amber) and those I had the honour to supervise in their own academic journey. I also want to thank some of my guides and supporters in the form of psychotherapists, thank you, Paula and Flori. I would really like to express my profound appreciation for my partner and friend, Mika! You've been with me for a big stretch of this journey, and you made me feel so supported, brave, smart, beautiful, but most of all, truly loved. I am grateful for all the places I have visited and lived in. For those amazing hours spent in nature, for the garden in which I spend most of the days in the pandemic, for the mountains, rivers and seas of Catalunya. Thank you again to my supervisors for being open and supportive of me to also live, as possible where I felt more at home.

Lastly, I also want to acknowledge myself. I acknowledge the opportunities and challenges which came with working with more Eastern and Western (Nordic) European ways and standards. While I had my own dose of self-deception at the start of this PhD journey, I am grateful for it, as without I would not have embarked on this adventure! Which was not easy and long, and sometimes not pleasant at all. Being here, in this moment, though, I am grateful I dared!

CHAPTER I. GENERAL INTRODUCTION AND THEORETICAL FRAMEWORK

1.1 Introduction and Motivation[1]

"At the most basic level, self-deception is fooling ourselves into believing something that is false – or – not believing something that is true."

Cortney S. Warren (psychotherapist)

On a popular Dutch television show, Belgian author Griet op de Beeck announced that while in therapy, she had discovered that her father had sexually abused her in childhood. She claimed that she never remembered the abuse until her therapist pointed out that her mental health symptoms were probably the result of repressed memories of childhood abuse. As a reaction to this story, other individuals came forward on social media with similar stories (Shaw & Vredeveldt, 2019). Repressed memories refer to the idea that autobiographic experiences (often traumatic) can be unconsciously blocked and remembered accurately, even years later (Battista et al., 2023; Dodier et al., 2024; Otgaar et al., 2019). This idea is at the center of a debate in legal psychology, as some of the recovered memories (memories remembered years later) may not reflect actual events, and several studies show the dangers of therapeutic suggestions on the creation of false memories, which are memories of not experienced events (McNally, 2003; Otgaar et al., 2019, 2022a). Cases such as the one presented above, due to their severe legal implications such as false accusations (Loftus, 2005; Shaw & Vredeveldt, 2019) have led to an increased interest in research on false beliefs and false memories (Howe & Knot, 2015; Kassin, 1997; Loftus, 2005; Otgaar et al., 2023; Wade et al., 2025). Consequently, using various paradigms, researchers showed that false autobiographical memories can be successfully elicited for past non-experienced events. These events varied from more common ("lost in the mall", Loftus & Prickel, 1995), to more traumatic ("tooth extracted by the dentist", Mazzoni & Memon, 2003), more implausible ("being abducted by UFOs, Otgaar et al., 2009), and events that allegedly occurred multiple times ("lost a toy", Calado et al., 2021; "swimming trucks fell off at the pool", Otgaar et al., 2022b).

Notably, in the literature on false memories, especially in the legal field, an important distinction has been made in the past two decades between recollection (the experience of mentally re-experiencing a past event) and belief (a metacognitive appraisal made regarding the

[1] This section contains adapted parts from the paper: Moldoveanu, M.G., & Visu-Petra, L. (2025). Self-deception beyond speculation: A narrative review of the empirical research on motivated false beliefs. *Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia*

genuine occurrence of an event in one's past) (Mazzoni et al., 2001; Mazzoni & Kirsch, 2002; Otgaar et al., 2025; Scoboria et al., 2004, 2014; Smeets et al., 2005).

The belief that repressed memories are related to unconscious mechanisms is still persistent and can have severe implications, such as leading to false beliefs and memories (Battista et al., 2023; Didier et al., 2024; Otgaar et al., 2019, 2022a). The idea of repression shares its conceptual roots with another construct and phenomenon, namely self-deception (Krickel, 2024; Tomaka et al., 1992). More specifically, self-deception and repression have been explained in psychoanalytic perspectives as related to or being defence mechanisms used to protect oneself from threatening information which allegedly happens out of our conscious awareness, in our unconsciousness. While repression was explained as a mechanism, self-deception was portrayed as a broader construct (Freud, 1960; Didier et al., 2024; Otgaar et al., 2019; Westland, S., & Shinebourne, 2009). Researchers' interest in self-deception was originally inspired by the psychoanalytic perspective, but self-deception is recognised as a complex phenomenon that is studied across a wide range of disciplines. Nonetheless, at the moment, there are significant differences among scholars regarding its precise definition and underlying elements, characteristics, and mechanisms (Bachkirova, 2016; Balcetis, 2008; Fan et al., 2022; Khalil, 2017; Mele, 1997; Paulhus & Holden, 2010; von Hippel & Trivers, 2011).

A proposed broad definition of self-deception, which includes several common elements from various perspectives, could be the process of acquiring and maintaining motivated false beliefs even in the face of contradicting evidence (Balcetis, 2008; Khalil, 2017; Mele, 2001; von Hippel & Trivers, 2011). Furthermore, a substantial part of the scientific literature on self-deception consists of theoretical and essentially speculative claims (e.g., Balcetis, 2008; Mele, 2001; Lauria et al., 2016; von Hippel & Trivers, 2011). These proposals focus on different ways of understanding the phenomenon or mechanisms which could facilitate self-deception, and they are not necessarily followed by empirical studies examining or testing the proposed models or hypotheses (but see von Hippel & Trivers, 2011; Smith et al., 2017). Studies on self-deception have yet to address fundamental aspects of the phenomenon, such as the systematic investigation of motivations or mechanisms that could lead to the successful formation of self-deception.

While there have been some advancements in understanding false beliefs and self-deception during the past decades, the paucity of concrete empirical evidence, especially in the case of self-deception, highlights significant gaps in understanding this phenomenon (Moldoveanu & Visu-Petra, 2025; Von Hippel & Trivers, 2011). Furthermore, in the case of self-deception the interchangeable use of terms like motivations (reason or reasons for acting or behaving in a

particular way), mechanisms (an established process by which something takes place or is brought about), or functions (purpose) and consequences of self-deception contributes to its limited conceptual validity despite its importance in everyday life. This is even more essential in legal contexts where false beliefs could lead to judicial errors (Ost, 2017; Howe & Knot, 2015).

1.2 Research Relevance

The practical relevance of studying false memories and false beliefs has been discussed in many studies in the past couple of decades (e.g., Bernstein et al., 2015; Howe & Knot, 2015; Loftus, 2005; Mazzoni et al., 2010; Scoboria et al., 2017; Wade et al., 2024) as an important aspect of the reliability of eyewitness testimony in the courtroom. The impact of self-deception has been studied predominantly in reaction to feedback (e.g., Fan et al., 2022; Liu et al., 2019, 2025), as well as the association with other psychological constructs such as subjective well-being (Dufner et al., 2019; Taylor et al., 2003). In the legal field, self-deception has been addressed in some studies looking at the validity of self-report measures relevant for judicial decision-making, such as measures of factors evaluated in the context of risk assessment (the likelihood that an individual will engage in harmful, criminal, or otherwise adverse behaviour in the future; Hildebrand et al., 2018). Self-deception could also be relevant in the case of symptom validity assessment, in which (usually intentional, malingering) under- or over-reporting of psychological symptoms is assessed to evaluate whether someone is faking (Boskovic et al., 2024; Merten & Merckelbach, 2013). Additionally, a current debate in the literature is about whether self-deception represents a response bias as it was originally proposed, or whether it manifests itself as a more stable disposition or trait (Mills & Kroner, 2005; Tan & Grace, 2008; Vigil-Colet et al., 2012).

Research on situational self-deception in the legal literature is lacking. We argue that self-deception might be identified in various facets of the legal field, for example, in internalised false confessions (an innocent person, genuinely coming to believe they are actually guilty) (Kassin, 1997). Another area in which self-deception might be present in the legal field could be in experts' tendency towards overconfidence (when individuals overestimate their own abilities, knowledge, or the accuracy of their judgments) and inflated estimations of abilities and decisions (Neal et al., 2018). Self-deception has been proposed to have evolved as a way to better deceive others (von Hippel & Trivers, 2011), and limited empirical evidence shows that indeed this might be the case, that self-deception is higher in deceptive contexts and could be associated with successful deception (Jian et al., 2019; Lu, 2012; Smith et al., 2017). Several

studies showed that lasting false beliefs could appear as residual effects of lying as well as faking psychological symptoms (Kunst & Winkel, 2015; Merckelbach & Merten, 2012; Otgaar et al., 2023; Riesthuis et al., 2023).

Focusing on false beliefs, and more specifically on self-deception, can allow a nuanced understanding of how false beliefs about one's past and present can form and persist, as well as the consequences of false beliefs in different settings. This introductory chapter describes the conceptual aspects and elements of false beliefs, with a more in-depth focus on a specific type of false belief, namely self-deception. Individual and contextual factors shaping false beliefs and self-deception are presented, alongside mechanisms, paradigms, and consequences of false beliefs and self-deception.

1.3 False Beliefs

1.3.1 Theoretical Accounts of False Beliefs

Several authors have suggested that we tend to accept appearances as reality, and we form beliefs regarding reality until they are proven otherwise. To update our beliefs, we use external information (e.g., from the exterior, through our senses such as hearing or seeing) and internal information (e.g., prior knowledge, our emotions, prior belief system, etc.; Seitz & Angel, 2022). Our beliefs can guide our behaviour, for example, if we believe that the person is innocent, we might try to find confirmatory evidence that they are indeed innocent. While beliefs are important as they sustain our understanding of our reality, beliefs are not always an actual representation of facts or reality. Additionally, as beliefs are internal representations and not direct reflections of reality, they can be inaccurate. When an individual is falsely convinced that an event or details from an event have occurred to them, we refer to this as false beliefs (Lampinen et al., 1998). One field addressing false beliefs is the legal psychology field, more specifically in false memory research (Otgaar et al., 2018a, 2018b, 2023; Scoboria et al., 2004, 2014; Smeets et al., 2005). Memory is generally reliable, but (most) researchers agree that memory does not record events like a video camera would. Instead, memory is a reconstructive process, meaning we rebuild our memories each time we recall them (Loftus, 2005). In some cases, due to different reasons, such as external influences (e.g., suggestion), false memories can be formed (Loftus & Pickrell, 1995).

1.3.2 Memory Errors

Memory can be affected by external (e.g., suggestion) and internal influences (e.g., lying) (Otgaar & Baker, 2018; Otgaar et al., 2023). These influences can lead to memory

errors, which are of details/events suggested as present, while they were not actually present, or of certain details/events that were not experienced when they were in fact experienced (Otgaar et al., 2023). These memory errors are referred to in the literature as omission errors (an experienced event is reported only partially or not reported at all; relevant information is forgotten or omitted) and commission errors (a non-experienced event or details are reported or experienced events or details are reported differently than experienced) (Loftus, 2005; Otgaar et al., 2023).

1.3.3 Distinction Between Recollection And Belief

Until the early 2000's when studying false memories, some researchers were measuring autobiographical belief, which is "all the autobiographical information that a person has, inaccurate, as well as accurate" (Scoboria et al., 2004, p.792), while others were focusing on reports and detailed accounts of events (Loftus, 1997). On the other hand, recollection refers to the experience one has of mentally reliving a past event, which is subjective (Scoboria et al., 2017). To date, multiple studies have shown that, for example, feedback and suggestions can impact the belief that an event occurred separately from the recollection of an event (e.g., Clark et al., 2012; Li et al., 2020; Mazzoni et al., 2010; Otgaar et al., 2014; Scoboria et al., 2017, but see also Li et al., 2020, 2022; Otgaar et al, 2016b, 2016c; Zhang et al., 2024; Wang et al., 2019a, 2019b). In light of these findings, beliefs and recollection have been proposed to be different constructs of memory (Mazzoni et al., 2001; Mazzoni & Kirsch, 2002; Scoboria et al., 2004, 2017). Research on false beliefs has shown that false beliefs can form or be inflated for non-experienced events (individuals come to believe suggested events actually happened to them, Scoboria et al., 2012) or be retracted (no longer believing an event has occurred after, for example, receiving feedback that the event did not happen, Otgaar et al., 2016b) or deflated (decreased) for experienced events (Otgaar et al., 2023; Polage, 2019). As with false memories, changes in beliefs can be elicited due to external influences (e.g., suggestion) or can occur spontaneously due to internal influences (e.g., lying) (Otgaar et al., 2023). Additionally, changes in belief might be influenced by several factors, such as contextual or individual factors (Riesthuis et al., 2023).

1.3.4 Individual and Contextual Factors Shaping False Beliefs

A range of individual and contextual factors which can influence the formation of both false memories and false beliefs have been studied, such as event plausibility, memory distrust, social feedback and emotional valence (e.g., Bernstein et al., 2009; Calado et al., 2021; Mazzoni et al., 2001; Otgaar et al, 2016a; Pezdek et al., 1997; Scoboria et al., 2006; Zhang et al., 2022,

2024). Additionally, executive functions (e.g., Shifting, Inhibition and Updating, Battista et al., 20201), cultural background (Wang et al., 2021), (Dark) Personality Traits (e.g., extraversion, neuroticism and psychoticism, Mirandola et al., 2020; psychopathy, Mirandola et al., 20232) have been addressed in some studies, although less systematically. Furthermore, most studies have examined these factors in isolation, despite the likelihood that multiple event characteristics co-occur in real-life situations. For example, Otgaar and colleagues (2022b) found that event frequency did not lead to differences in false memory formation and argued that this might be influenced by the choice of event, as more or less plausible events, in some studies, were shown to influence the rate of false belief and memory formation (Mazzoni et al., 2001; Scoboria et al., 2012, but see also Strange et al., 2006). Another individual differences factor studied and which was shown to be practically relevant when discussing autobiographical belief formation, is memory distrust, the subjective appraisal of one's memory functioning (Zhang et al., 2022, 2022, 2024b).

Motivation has also emerged as a relevant factor (Scoboria et al., 2014), meaning a motivation to not remember the event, with 4.3% of participants mentioning this as a motive for retracting belief in the event in college students. Similarly, Polage (2012) found an association between the fabrication inflation effect (inflated beliefs in a previously lied-about event) and levels of dissociative experiences, providing preliminary evidence for a motivated account of false belief formation that will be discussed in more depth in the next section on self-deception.

1.3.5 Consequences of False Beliefs

While in real-world cases, false memories and false beliefs can lead to severe consequences (Newman et al., 2022; Otgaar et al., 2022b), there have also been experimental studies addressing some consequences of false beliefs on, for example, decision-making (Wang et al., 2019b). False beliefs have been shown to contribute to memory errors, as individuals may misremember events in ways that support their inaccurate beliefs (Loftus, 2005; Bernstein et al., 2015). Research has shown that false beliefs in the absence of recollection led to negative consequences in attitudes and behaviour (e.g., food preferences, Bernstein et al., 2015; Geraerts et al., 2008; Pezdek & Freyd, 2009). Wang et al. (2019a, 2019b) showed that false beliefs undermine problem-solving, decision making and behaviour (food preferences) and that false memories and false beliefs are increased when one processes information related to oneself (self-referencing) compared with information related to others (Wang et al., 2024). Moreover, the consequences of acquiring false beliefs are not always negative, with, for example, a study

by Laney et al. (2008) showing positive false memories could similarly enhance food preferences by inducing positive attitudes towards asparagus and implanting the false suggestion that participants had enjoyed asparagus the first time they tried it.

1.4 Self-deception

1.4.1 Perspectives on Self-Deception[2]

Philosophical approaches to self-deception focus on defining why self-deception happens in individuals and the minimal necessary conditions for self-deception to occur (Gur & Sackeim, 1979; Mele, 2001). The two most influential views in philosophy regarding the conceptualisation of self-deception were postulated: the deflationary (non-internationalist) and the inflationist (internationalist) views. In the *inflationist* view, the self is divided into distinct parts that could be in conflict (one "part" holds a belief, while the other holds a contradicting belief). For example, some theories of self-deception were based on the idea of the split between the conscious and unconscious (Jian et al., 2019; von Hippel & Trivers, 2011). While the existence of a structural partition of the mind is still postulated (e.g., Schwardmann & van der Weele, 2019), these perspectives rather describe self-deception as related to or biases in information processing processes (i.e., information gathering; Smith et al., 2017; information processing and memory processes, von Hippel & Trivers, 2011) and belief formation (Mijović-Prelec & Prelec, 2010). Conversely, in the *deflationary* view proposed by Mele (1997, 2000, 2001), it is argued that self-deception is a general category of motivated-biased judgment (Mele, 1997). This view proposes that self-deception is not intentional, namely, the intention to deceive oneself is not present as in interpersonal deception (DePaulo et al., 1996), but what drives self-deception are motivational factors (e.g., emotions, desires).

Deriving from early psychoanalytical and philosophical works, the psychological literature on self-deception has been divided for decades between two perspectives regarding the motives and mechanisms of self-deception. One of these is an *intrapersonal* perspective in which the self-deceivers are deceiving the self for their own benefit, which is to protect the self from threatening information (Freud, 1946; Mele, 2000; Sackeim & Gur, 1979; a concept similar to the literature on psychoanalytic denial and repression). On the opposite end is the *interpersonal perspective* in which the self-deceiver is proposed to be deceiving the self to

[2]This section contains adapted parts from the paper: Moldoveanu, M.G., & Visu-Petra, L. (in press). Self-deception beyond speculation: A narrative review of the empirical research on motivated false beliefs. *Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia.*

better deceive others (for example, being a byproduct of other-deception; Schwardmann & van der Weele, 2019; von Hippel & Trivers, 2011).

Furthermore, over the years, self-deception was mostly described in the psychological literature either as a negative (yet adaptive) or defensive phenomenon (a defence mechanism; such as denial, or repression, Johnson, 1995; self-protection, Alicke, & Sedikides, 2009; Greve & Wentura, 2010) or as a positive, more offensive one, focusing on enhancing and attributing positive and desirable traits to oneself, often associated involved in positive illusions and optimism (Robinson, et al., 2009; Taylor, 2003). Additionally, self-deception has been discussed as a process (describing the mechanisms of how self-deception takes place within the individual, e.g., motivated cognition, Balcetis, 2008; von Hippel & Trivers, 2011) as well as a product (e.g., false self-beliefs which are stated; Balcetis, 2008; Mijović-Prelec & Prelec, 2010). Beyond theoretical definitions, psychological literature focused for the first time on operationalising self-deception for conducting empirical studies (e.g., Sackeim & Gur, 1979).

1.4.2 Paradigms in Self-Deception Research[3]

Empirical research on self-deception in psychology has taken two main approaches: examining self-deception as a *trait, disposition,* or *response bias* (mostly correlational studies; Paulhus, 1991) or investigating it as a *situational behaviour* through experimental paradigms (Chance et al., 2011; Gur & Sackeim, 1979; Mei et al., 2023). Gur and Sackeim (1979) conducted one of the first experiments in the field to elicit self-deception and prove its existence. Another early pioneering attempt at eliciting self-deception was made by Quattrone and Tversky (1984). These early attempts at eliciting self-deception suggested the possibility of contradictory beliefs associated with some motivations (e.g., reduced cognitive discrepancy, belief-consistent information processing). While these paradigms focused mainly on reactions to past behaviours (retrospective), more recent paradigms elicit self-deception using future, hypothetical situations (e.g., forward-looking paradigm, Chance et al., 2011).

Proposed as a better attempt to measure the false beliefs necessary to show that self-deception has been elicited, the *forward-thinking paradigm* involves participants taking a test, with half being given access to the answers (e.g., answers are included at the end of the test), thus offering a possibility to cheat. After the initial test, participants are asked to predict their future performance in a hypothetical test without answers. In some studies, a second test is also

[3] This section contains adapted parts from the paper: Moldoveanu, M.G., & Visu-Petra, L. (in press). Self-deception beyond speculation: A narrative review of the empirical research on motivated false beliefs. *Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia.*

given to calculate the differences between future estimations and actual performances in a second test. However, the paradigm is focused on the differences between the estimations of a group that did not have access to the answers compared with a group that did have access. The typical finding using this paradigm is an overestimation of the (hypothetical) future performance, interpreted as self-deception, with participants who had access to answers in the first test predicting significantly higher scores for the second test compared to the group without access to the answers, despite both groups knowing they do not have the answers for the second test (Chance et al., 2011, 2015; Fan et al., 2022; Jian et al., 2019; Liu et al., 2019; Mei et al., 2023).

When approached as a *trait, disposition,* or *response bias*, self-deception has been mostly studied using the rationale that people tend to deny negative qualities about the self and enhance positive ones (Sackeim & Gur, 1979; Paulhus & Reid, 1991). Sackeim & Gur (1979) developed the first questionnaire to measure self-deception (Self-Deception Questionnaire, SDQ), which was later modified and further validated by Paulhus and Reid (1991) as the Balanced Inventory of Desirable Responding (BIDR, Paulhus, 1991; later known as Paulhus Deception Scales, PDS, Paulhus, 1998). Some items of these measures are: “I am a completely rational person.”, “I never regret my decisions.”, “I am fully in control of my emotions.”, “I am a very disciplined person.” (Paulhus, 1998). In empirical studies, self-deceptive enhancement has been used especially when using self-reported measures and assessments as a way of increasing the validity of measurements (Vigil-Colet et al., 2012) and as part of the construct of social desirability and the proposed tendency to dissimulate in more legal contexts (Crowne & Marlowe, 1960; Hildebrand et al., 2018; Paulhus, 1991). Some studies looked at the association of self-enhancement with other constructs such as personal adjustment (Dufner et al, 2019; Sheridan et al., 2015; Taylor et al., 2003), anxiety (Sheridan et al., 2015), or personality (e.g., Dark Triad, Wright et al., 2012). However, there is a debate in the scientific literature on how to best approach self-deceptive enhancement tendencies: as a response style, a personality trait, or a disposition (Barry et al., 2016). While there are some proponents of treating self-deceptive enhancement more as a disposition or personality trait (Mills & Kroner, 2005; Vigil-Colet et al., 2012) which adds unique variance to some phenomena (e.g., violent recidivism, Mills & Kroner, 2005), many correlational studies include self-deception as a potential confounding variable operationalising self-enhancement as a response bias, particularly when measuring constructs sensitive (or less sensitive) to self-presentation effects such as aggression or dynamic risk factors (for discussions regarding this see also Hildebrand, et al., 2018; Vigil-Colet et al., 2012).

1.4.3 Individual and Contextual Factors Shaping Self-Deception

Some of the individual factors addressed so far are personality traits and individual factors such as self-esteem, as well as more cognitive and affective factors such as cognitive inhibition, locus of control, and self-efficacy as a component of self-regulation (Erez et al., 1995; Paulhus, 1991; Sengupta et al., 2020; Wright et al., 2021). Dispositional self-deception (self-deceptive enhancement) was found to be associated with Bright Personality Traits such as Extraversion, Openness, as well as Agreeableness and Conscientiousness (Berry et al., 2007; Paulhus, 1998; Lu & Chung, 2014). Furthermore, dispositional self-deception was associated with Dark Personality Traits, especially Narcissism (Paulhus & Holden, 2009). The most consistent finding in the literature on self-deception is that increased contextual ambiguity (e.g., general feedback which can be easily interpreted in a self-serving way) leads to an increase in self-deception (Johnson, 1995; Ren et al., 2019; Sloman et al., 2010; Mei et al., 2023). Moreover, some of these studies (Ren et al., 2019; Mei et al, 2022) also focused on offering feedback and showed that negative feedback led to a decrease in self-deception compared with more neutral or positive feedback (see also Chance et al., 2015). Another factor which has scarcely been addressed is the differences in social status (Lu & Chang, 2014; Jian et al., 2019). Additionally, limited studies have addressed cognitive load with preliminary findings that it is reduced compared with deception (Jian et al., 2019). Self-deception has also been conceptualised as a form of affective coping, where individuals unconsciously manipulate their beliefs or memories to avoid distressing emotions and maintain psychological comfort (Lauria et al., 2016). Thus, emotion regulation was postulated as a factor which influences and plays a significant role in self-deception. Yet, research is scarce on self-deception and emotion regulation and affective factors (Arndt et al., 2013).

While individual factors are important, contextual factors could also influence self-deception. For example, one such factor is culture. While the effect of culture has not been addressed for situational self-deception, dispositional self-deception or self-deceptive enhancement was found to be sensitive to culture (Sedikides et al. 2003, 2005). Another important aspect, especially for dispositional self-deception, is the context or setting in which the measure is being taken. Similar to culture, the influence of the situational self-deception context or setting has been less investigated (rather, context characteristics such as ambiguity have been addressed). In settings in which individuals might want to avoid more serious punishments or gain benefits (e.g., insurance fraud), self-deception might be increased (e.g., Merckelbach & Merten, 2012); however, this has not yet been investigated.

Lastly, but crucial for self-deception, is the motivation for engaging in self-deception, which could be seen as an individual factor. Motivations for engaging in self-deception are mentioned in many theoretical works as one of the main distinctive features of self-deception, compared to merely acquiring false beliefs (Mele, 1998; von Hippel & Trivers, 2011). Using paradigms for situational and dispositional or trait self-deception, empirical research has been conducted which investigates the role of internal and external motivations in self-deception. However, studies are scattered and often treat motivations separately, making it difficult to have an overview.

1.4.3.1 Motivations Shaping Self-Deception: Empirical Evidence[4]

The proposed motivations for engaging in self-deception can be split into two types or categories: internal or self-oriented (mirroring the intrapersonal perspectives on self-deception) and external or other-oriented (mirroring the interpersonal perspectives on self-deception). The internal motivations are given by one's internal state (e.g., behavioural tendencies of approaching good and avoiding bad; Tice & Masicampo, 2008), and the external motivations refer to external gains and benefits (e.g., avoiding punishments by deceiving others successfully; von Hippel & Trivers, 2011). While internal and external influences are frequently discussed in research about lying to others (e.g., DePaulo et al., 1996; Otgaar et al., 2023), these categories and relevant empirical data have not been presented, to our knowledge, in a single, comprehensive review regarding self-deception. Furthermore, empirical studies investigating motivations for self-deception have often focused on a single motivation or perspective, such as deceiving others (e.g., Lu et al., 2014), or focused mainly on external motivations (e.g., financial incentives, social recognition, Chance et al., 2011). Including internal motivations is highly relevant as their impact and association with self-deception may differ. For instance, Hirschfeld et al. (2008) studied 429 college students and found that trait self-deception was positively associated with intrinsic motivations but negatively associated with extrinsic motivations. This suggests that individuals with higher self-deceptive tendencies report being more motivated by internal goals than external factors. Fewer studies explored the interplay between external and internal motivations, suggesting their interdependence and mutual influence (e.g., increased confidence and financial motives; Mijovic-Prejelec & Prelec, 2010). See Table 1 for an overview of the proposed motivations in psychological research.

Table 1

[4] [4] This section contains adapted parts from the paper: Moldoveanu, M.G., & Visu-Petra, L. (2025). Self-deception beyond speculation: A narrative review of the empirical research on motivated false beliefs. *Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia.*

External and Internal Motivations for Self-Deception

External Motivations	Internal Motivations
Gaining benefits	Increased confidence in oneself
Social validation and/or social recognition	Decrease in negative short-term affect
Deceiving others better	Reduced cognitive discrepancy or dissonance
Gaining financial incentives	Increased subjective well-being
Avoiding punishments	Increase in short-term positive affect
	Maintenance of (moral) self-concept
	Approaching good and avoiding bad (behavioural tendencies)
	Reduced cognitive load
	Avoiding threatening information

1.4.3.1.1 External Motivations.

Some of the most iterated external motivations for self-deception studied or proposed in the literature are gaining benefits or avoiding punishments (e.g., punishment for deception, von Hippel & Trivers, 2011; financial incentives, Mijović-Prelec & Prelec, 2010), social validation and/or social recognition (Baumeister & Cains, 1992; Chance et al., 2011; Dufner et al., 2019; Lamba & Nityananda, 2014), and deceiving others (Lu, 2012; Smith et al., 2017; von Hippel & Trivers, 2011).

1.4.3.1.2 Internal Motivations

Internal motivations for self-deception proposed to date and studied in the literature are increased confidence in oneself (Mijović-Prelec & Prelec, 2010), a decrease in negative short-term affect, cognitive discrepancy (Sackeim & Gur, 1979), or cognitive dissonance (Merchelbach & Merten, 2008; Otgaar et al., 2023), increased subjective well-being (Baumeister & Cains, 1992) and increase in short-term positive affect, the maintenance of (moral) self-concept (Baumeister & Cains, 1992; Lu & Chang, 2011; Mazar et al., 2008), behavioural tendencies of approaching good and avoiding bad (Tice & Masicampo, 2008), and reduced cognitive load (Butterworth et al., 2022; Jian et a., 2019; von Hippel & Trivers, 2011). While proposals such as approaching good and avoiding bad have not yet been tested (Tice & Masicampo, 2008), other motivations have been investigated in some empirical studies. Below, we will explain each motivation and review some empirical studies that addressed them.

(Moral) Self-concept Maintenance. The idea of self-deception to preserve one's (positive) moral self-concept has been extremely influential. Baumeister (1996), one of the first proponents of self-deception being motivated by the maintaining of well-being, argued that people are strongly motivated to maintain favourable self-concepts about themselves and when

these self-concepts are threatened, people sometimes resort to self-deceptive strategies (i.e., deceiving themselves so they do not have to update their self-concept; Baumeister, 1996; Baumeister & Cairns, 1992; see also Greenwald, 1980, Sedikides and Skowronski, 1997). Only a few studies tried to test the self-concept maintenance theory in relation to self-deception, mostly by looking at the way people react to receiving correcting information (e.g., feedback, Baumeister and Cairns, 1992, see also Weinberger et al., 1979). However, in this study, moral self-concept was not directly measured, nor was self-deception; thus, the influence of moral self-concept maintenance on self-deception is still unknown. In another study investigating self-reported trait self-deception in association with measured trait moral self-concept, Lu and Chang (2011) showed that dispositional self-deception was associated with a positive moral self-concept in a sample of college students.

1.4.4 Consequences of Self-Deception

The literature on self-deception has been investigating the consequences of self-deception through experimental and correlational studies that looked at the short-term impact on everyday functioning (Johnson et al., 1995; Peterson et al., 2003; Chance et al., 2011, 2015) and also the long-term impact on psychological adjustment and interpersonal functioning (Taylor et al., 2003; Dufner et al., 2019). Overall, empirical evidence shows that self-deception might have negative consequences in the short term (distorted self-views, Chance et al., 2011, 2015; failure to integrate feedback, Johnson et al., 1995, 1997; Peterson et al., 2003). For example, Johnson et al. (1997) looked at the influence of dispositional self-deception (denial and enhancement) in treating failure on performance and emotional response (e.g., anger and hostility). Greater enhancement predicted significantly worse problem-solving and hostility, and greater denial predicted significantly worse post-failure problem-solving. In the case of dispositional self-deception, it was initially seen as potentially beneficial for mental health as it was consistently positively associated with measures of psychological adjustment (Sedikides & Gregg, 2008; Taylor et al., 2003) and negatively associated with anxiety and depression (Arndt et al., 2013; Sheridan et al., 2015). However, some studies have revealed mixed outcomes. For example, self-enhancers may make positive first impressions, but in the long term, are evaluated by others as less trustworthy, as positive interpersonal perceptions diminish over time (Dufner et al., 2019; Paulhus, 1998). Moreover, in the legal field, the elicitation of self-deception in contexts involving eyewitness testimony may lead to negative consequences for the accuracy of event memory (Merckelbach & Merten, 2012). Additionally, the dispositional aspect of self-deception is a significant factor in self-report measures that inform critical decisions, such as

risk assessments (Hildebrand et al., 2018). In the following subsections, we will explore these areas more in-depth.

1.4.4.1 Memory Effects of Self-deception

Intentional deception and self-generated lies have been shown to lead to memory errors (Otgaar et al., 2023; Riesthuis et al., 2022). However, to our knowledge, no empirical studies have directly examined the effects of self-deception on memory (but see below Jian et al., 2019; Lu & Chang, 2014). Although previous sections focused on memory errors, numerous studies have examined self-serving memory bias, a phenomenon closely related to self-deception, particularly in its dispositional dimension, drawing on the well-established literature on self-enhancing or self-serving bias (Miller & Ross, 1975; see also Shepperd et al., 2008). The self-enhancing or self-serving bias refers to the fact that individuals tend to overestimate their positive traits (e.g., intelligence, capacity, morality, Taylor et al., 2003) and downplay what they perceive as undesirable present or future qualities (*self-serving bias,* Dufner et al., 2019, but see also the self-verification theory, Swann, 2012). The self-serving memory bias can be seen, for example, when recollecting daily experiences, people better recall their good behaviour than their bad behaviour (D'Argembeau & Van der Linden, 2008). While the self-serving bias is referred to in studies which investigate self-deception (either as a disposition or situational; Taylor et al., 2003), research exploring the association between self-deception and self-serving memory bias is scarce.

1.4.4.2 (Un)Desirable Behaviours and Attitudes – Under and Over-Reporting

Self-deception has been widely studied as a component of social desirability across psychological domains, including legal, clinical, and organisational, where people often deny negative traits and inflate positive ones (Paulhus & Reid, 1991; Perinelli & Gremigni, 2016; Sackeim & Gur, 1979; Tan & Grace, 2008; Zerbe & Paulhus, 1987). Several measures have been developed to assess social desirability, with the most used being the Marlowe Social Desirability Scale (Crowne & Marlowe, 1960), which measures social desirability as a unidimensional construct. Another validated scale is the Balanced Inventory of Social Desirability Responding (Paulhus, 1991; later known as Paulhus Deception Scales, PDS, Paulhus, 1998), which treats social desirability as two factors: one oriented towards the self, self-deceptive enhancement, and a factor oriented towards others: impression management (see also Section 1.4.3). Some measures of psychological correlates, such as personality, include a subscale oriented towards dissimulation and social desirability as part of the measure (e.g., the Lie Scale used to detect intentional underreporting of psychological symptoms in the Minnesota

Multiphasic Personality Inventory, MMPI, Tellegen & Ben-Porath, 2008). Either stand-alone or integrated social desirability measures are used to verify the validity of self-reported measures, which, for example, in the legal field, are essential in certain proceedings (e.g., risk assessment, Hildebrand et al., 2018). Various approaches have been developed to verify or reduce validity threats to self-report measures (for a review, see Tan & Grace, 2008).

In the general population, dispositional self-deception has been associated with measures of personal adjustment (Dufner et al, 2019; Sheridan et al., 2015; Taylor et al., 2003). In legal settings, however, there is a proposed tendency that individuals tend to dissimulate in their self-report measures to influence judicial decision-making and, for example, avoid punishment (Crowne & Marlowe, 1960; Hildebrand et al., 2018; Paulhus, 1991). No studies investigated different contexts comparatively, such as community settings and forensically relevant settings (e.g., offenders). Nonetheless, self-deceptive enhancement has been found to correlate with measures relevant to legal contexts, both in the general population and in forensic settings. These include factors such as anger, criminal recidivism, antisocial attitudes, aggression, and deception (Barry et al., 2016; Buta et al., 2021; DePaulo et al., 1996; Kolarcik et al., 2016; Mathie & Wakeling, 2011; McEwan et al., 2009; Mills et al., 2003; Mills & Kroner, 2005, 2006; Visu-Petra et al., 2014). More recently, scholars pointed out that these associations may arise because self-deception can function either as a response style, as originally considered, or self-deception in this case could be rather dispositional or a stable personality trait. Some scholars argue that self-deceptive enhancement contributes unique variance to specific outcomes, such as violent recidivism (Mills & Kroner, 2005) and should therefore be considered a personality trait (see Barry et al., 2016; Vigil-Colet et al., 2012).

1.5 Theoretical Frameworks: False Beliefs and Self-Deception

1.5.1 Theoretical Models of False Beliefs

One of the first influential models which emerged was the Nested Model of Belief and Memory (Scoboria et al., 2004), which posits that for a false memory to be formed, first an event has to be generally plausible, personally plausible and believed to have occurred; only after the event could be remembered. Furthermore, while false memories and false beliefs could be elicited separately, false memories could also be elicited in the absence of belief (Mazzoni et al., 2010). Following the finding regarding the separation between belief and recollection (Clark et al., 2012; Mazzoni et al., 2014; Otgaar et al., 2013; 2017), the Dissociated Model of Belief and Recollection was proposed by Scoboria et al. (2014). In this model, belief and

recollection are proposed as two independent and continuous components of memory, leading to events that can be high or low on autobiographical belief and high or low on recollection.

1.5.2 Theoretical Models of Self-Deception

Self-deception as a psychological construct is still undergoing a process of theoretical and methodological clarification. Several models have been proposed in the literature that try to address essential elements and characteristics of self-deception, such as motivation and mechanisms through which self-deception is achieved (Lauria et al., 2018; Mele, 2001; von Hippel & Trivers, 2011). Additionally, empirical studies also focus on individual and contextual factors influencing self-deception, as well as the consequences in everyday situations of engaging in self-deception (e.g., Chance et al., 2011, 2015; Mei et al., 2023; Ren et al., 2019).

Some of the most influential models on self-deception come from philosophy and evolutionary biology and are proposed by Freud (1960), Mele (1998, 1999, 2000, 2001; see also Balcetis, 2008) and von Hippel & Trivers (2011; see also Trivers, 2006) and Sackeim & Gur (1979; see also Paulhus & Reid, 1991). The psychoanalytic model (Freud, 1960) proposes that self-deception is a negative and adaptive defence mechanism that operates unconsciously to protect the individual from anxiety, guilt, or psychological distress arising from unacceptable thoughts, desires, or impulses. Mele (1998) and Balcetis (2008) proposed that self-deception acts as a motivation-based judgment. Von Hippel & Trivers (2011) proposed that self-deception evolved as a byproduct of interpersonal deception, and it is defined as a biased information processing in which an underlying motivation is out of awareness. They define self-deception as a form of biased information processing in which motivational factors operate outside of conscious awareness, leading individuals to selectively attend to, interpret, or recall information in a way that supports desired beliefs or self-views. As a result, empirical studies using this perspective on self-deception have been studying it as the maintenance of positive, inflated beliefs about oneself (Chance et al., 2011, 2015). While the previous models treated self-deception as more of a situational event, Sackheim & Gur (1979) and later developed by Paulhus & Reid (1991), proposed that self-deception is a component of social desirability. Initially, self-deception was proposed to be a response bias that influences self-report measures. Paulhus and Reid (1991) proposed two forms: self-deceptive denial and enhancement. Later studies showed that self-deceptive enhancement is associated with over-reporting and under-reporting, with differences in different settings (e.g., Mills & Kroner, 2005; Vigil-Colet et al., 2012; Tan & Grace, 2008). Irrespective of the approach, one common element to most of the theoretical work on self-deception in psychology is the mention of specific motivations which

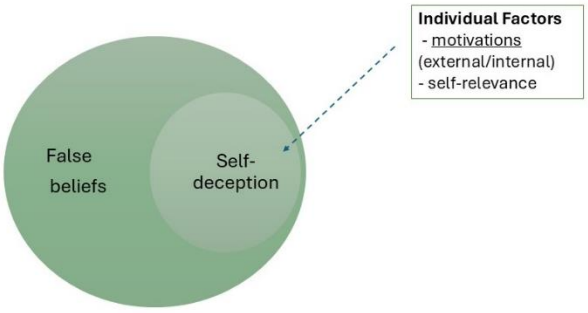
lead to the formation or increase/decrease of this phenomenon (Balcetis et al., 2008; Mele, 2001; von Hippel & Trivers, 2011).

1.6 Proposed Integrative Model of (Motivated) False Beliefs – Individual and Contextual Influences

Building upon various previous models of false beliefs and self-deception, as well as from empirical research and theoretical work (Balcetis, 2008; Mele, 200; Scoboria et al., 2004; 2017; Von Hippel & Trivers, 2011), we propose that false beliefs arise during information processing and that self-deception is a distinct case of false beliefs. More specifically, both false beliefs and self-deception arise from errors in information processing, with cases of self-deception being elicited due to specific motivations (see Section 1.4.4 for a detailed explanation and Figure 3 for a visual representation).

Figure 3

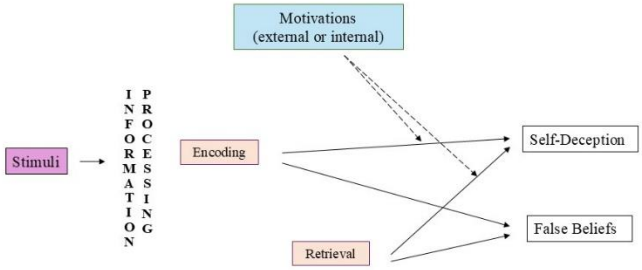
Proposed Relation Between False Belief And Self-Deception



Furthermore, we distinguished between false beliefs for non-experienced events (falsely believing that something did happen; in self-deception, fabricated or enhanced information) or details and false beliefs for experienced details/events (falsely believing that something did not happen; in self-deception, rejected or denied information) for both memory errors of belief, with or without specific motivations (Mele, 2001; Otgaar et al., 2023; Paulhus & Reid, 1991). Moreover, false beliefs and self-deception could occur as a result of interpersonal (external) or intrapersonal (internal) influences (Otgaar et al., 2023; Mijović-Prelec & Prelec, 2010); however, in the case of self-deception, these influences are related to motivations that are not intentional or the primary goal of the individual, but rather it is a case of **motivated reasoning** (Balcetis, 2008). Thus, we propose that both false beliefs and self-deception are cases of information processing errors, with self-deception representing a motivated error or distortion (see Figure 4).

Figure 4

Proposed Model of False Belief and Self-Deception as Information Processing Errors



*Note***.** When new information is encoded or retrieved, errors in processing this information might appear, such as the formation of false beliefs. When there is information processing is influenced by external or internal motivations, the errors represent motivated false beliefs (self-deception); however, this distinction has not been systematically tested, thus it is represented with a dotted line.

While the literature on false belief in the legal field mostly focuses on false beliefs regarding specific events or details of an event, false beliefs can also be developed regarding more general about one's abilities, behaviours or attitudes. While for false beliefs, research in the legal system has been investigating more situational false beliefs, in self-deception, another important distinction in self-deception is its dispositional or trait character (Paulhus & Reid, 1991; Paulhus, 1998; Virgil-Colet et al., 2012). Using these four dimensions: type of event of information, with or without motivation, source of influence, and type of self-deception, we propose the following model of forms (motivated) false beliefs presented in Figure 5. This model distinguishes between false beliefs and self-deception based on whether the error is motivated. False beliefs may arise following encoding or retrieval errors, while self-deception involves motivated reasoning. Both can refer to experienced or non-experienced events or event details and be shaped by internal (intrapersonal) or external (interpersonal) influences.

Figure 6

Proposed Forms of (Motivated) False Beliefs

	Non-Motivated (False Beliefs/Errors)	Motivated (Self-Deception*)
Experienced event/event details	False beliefs about denied real experiences (e.g., "I didn't do that")	Rejected/Denied (e.g., "I did not cheat", therefore "I am an honest person")
Non-experienced event/event details	False beliefs about fabricated events (e.g., "That happened" when it didn't)	Fabrication/ Enhancement (e.g., "I know everything in my field")

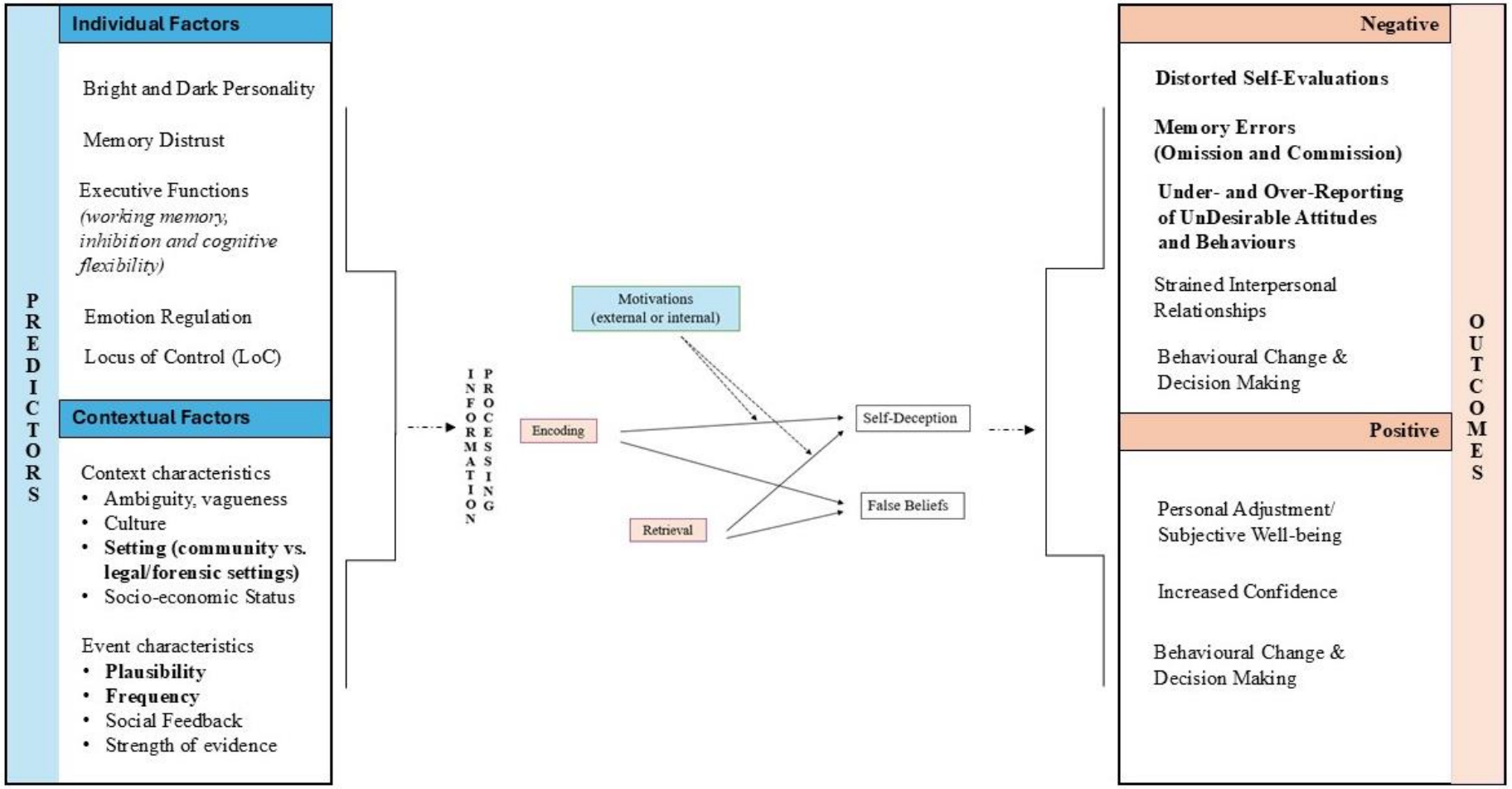
Note. Both non-motivated and motivated false beliefs can be elicited due to external (e.g., suggestions) or internal influences (e.g., lying). While false beliefs can have a more enduring character, self-deception has been characterised as manifesting in specific situations and at the trait level characteristic (e.g., dispositional).

Several individual and contextual factors have been identified as influencing both false beliefs and self-deception, with several studied in both domains. Lastly, false beliefs and self-deception could have positive and negative consequences in everyday situations, as well as in other settings, such as the legal domain. To date, no existing model has comprehensively integrated false beliefs theories and empirical evidence with self-deception defined as motivated false beliefs, while also accounting for the various influencing factors and potential consequences of these phenomena. Thus, based on the theoretical models and empirical evidence presented in this introductory chapter, we propose a new integrative model that differentiates between false beliefs and self-deception based on motivation and identifies a range of individual and contextual factors that shape both false beliefs and self-deception within the broader framework of motivated false beliefs (see Figure 6).

As there is an overlap in the conceptual understanding of false belief and self-deception, the literature on both could benefit from addressing these in parallel or in studies which combine both constructs. Additionally, our model outlines positive and negative consequences that may arise from forming false beliefs or engaging in self-deception. While some of the proposed factors as well as consequences may pertain more to either false beliefs or self-deception, this model can be used to systematically address false beliefs and self-deception in everyday settings, as well as in legal and more forensically relevant contexts. In the current thesis, several individual and contextual factors and consequences of self-deception on memory and under- and over-reporting are addressed.

Figure 6

Proposed Integrative Model of (Motivated) False Beliefs – Individual and Contextual Factors Shaping False Beliefs and Self-Deception



Note. Factors and outcomes, along with corresponding motivations addressed in the current thesis, are indicated in bold.

2 CHAPTER II. RESEARCH AIMS AND METHODOLOGY

2.1 Theoretical Aims

The current thesis focuses on individual and contextual factors which shape the formation of false beliefs and self-deception as motivated false beliefs, as well as the consequences of engaging in self-deception in different settings. To address this overarching aim, several sub-aims were considered.

Aim 1- The first aim was to elicit and assess **various types of false beliefs and self-deception**. Specifically, we aimed to examine the following ones situational false beliefs and self-deception that occur in everyday life, false beliefs and self-deception that are experimentally elicited in controlled settings, and dispositional self-deception as assessed in applied contexts, such as among violent offenders currently serving their sentences.

Aim 2- The second aim was to **identify individual and contextual factors shaping false beliefs and self-deception**.

Aim 3- The third aim of the current thesis was to investigate **different settings or contexts in which false beliefs and self-deception are formed**.

Aim 4- The fourth aim was to **explore negative (and positive) consequences of self-deception**.

2.2 Methodological and Practical Aims

While studying false beliefs has its own set of challenges, studying self-deception comes with a variety of methodological challenges affecting both experimental designs as well as interpretations of findings (Mijovic-Prelec & Prelec, 2010). First of all, the concept itself still has a high degree of conceptual ambiguity, and it is difficult to distinguish it from other phenomena such as overconfidence or other lying (Khalil, 2017; von Hippel & Trivers, 2011). For example, in experimental paradigms which induce self-deception (e.g., the forward-thinking paradigm), self-deception is inferred from the discrepancy between predictions and actual performance. However, it is challenging to ensure that the predicted false beliefs about future abilities are self-deceptive and not other deception or simply non-motivated errors or false beliefs. Furthermore, its definition and situational vs response bias, disposition or trait gives an added nuance to self-deception (Paulus & Reid, 1991) as few studies have addressed both in studies (Chance et al., 2011, 2015), thus making it seem as if there are two different constructs.

Aim 1- The first methodological aim was **to develop and extend different methodologies** for studying false beliefs, especially self-deception (e.g., mixed-methods surveys, blind implantation paradigms, forward-thinking paradigm) to capture self-deception and false beliefs more reliably.

Aim 2 - One main debate in the literature on self-deception as a facet of social desirability is whether it represents a response bias or a more stable disposition or trait-like feature which adds variance to self-reported measures of attitudes and behaviours. The second methodological aim for this thesis was t**o explore the predictive value of standardised measures of dispositional self-deception across different settings** (e.g., Paulhus Deception Scales, BIS/BAS, FFM) in identifying individuals prone to aggression and deception (using as a proxy an attitude towards deception validated measure).

Aim 3 - The third aim of this thesis was **to integrate data collection from experimental and real-life contexts**, including general population samples and incarcerated populations, to enhance ecological validity.

Aim 4 - Lastly, the final aim was more **practical and applied** and was **to highlight the legal and forensic relevance of false beliefs and self-deception.**

3 CHAPTER III. ORIGINAL RESEARCH CONTRIBUTIONS

3.1 Study 1: A mixed-methods survey on self-deception[5]

"When A deceives A (i.e., himself) into believing that p is true, he knows or truly believes that p is false while causing himself to believe that p is true. Thus, A must simultaneously believe that p is false and believe that p is true. But how is this possible?"

(Mele, 1998)

Self-deception has been gaining attention in *psychology* since the early psychoanalytic writings of Freud (1960). He described it as a defence mechanism, which is negative yet adaptive, enabling individuals to avoid threatening information through other defence mechanisms such as denial and rationalisation. Moreover, self-deception is similar to repression (the mechanism through which threatening information is automatically and unconsciously removed from conscious awareness, see Otgaar et al., 2019). According to the psychoanalytic view, self-deception also occurs by storing the threatening information in the unconscious, making it inaccessible to conscious awareness. Early psychological research on self-deception, rooted in psychotherapy case studies, posited that self-deception was widespread (Freud, 1991; Westland & Shinebourne, 2009); however, it did not explicitly mention its prevalence. Examples of self-deception were not provided, as the authors relied on anecdotal examples to protect patient confidentiality. Although self-deception is often described as a widespread phenomenon (von Hippel & Trivers, 2011; Lauria et al., 2016), its prevalence and frequency remain unexplored. To address this gap, the present study aimed to investigate the *prevalence* and *frequency* of self-deception, collect *real-life examples* and examine its *characteristics* as discussed in the literature, focusing on a subsample of emerging adults from the general population as a first step.

The questions and choice of characteristics and elements of self-deception were informed by previous theoretical and empirical work on self-deception (Baumeister, 1993; Chance et al., 2011; Gur & Sackeim, 1979; Lauria et al., 2018; Mele, 2001; Trivers, 2000; Sackeim & Gur, 1979; von Hippel & Trivers, 2011, etc.). These were the following: intentionality and awareness of self-deception, motivations, type of motivations: interpersonal or intrapersonal, type of self-deception: situational or dispositional, awareness of contradictory evidence and strategies for maintaining self-deception, retraction, emotions and affect,

[5] The content of this subchapter is a manuscript accepted for publication: Moldoveanu, M.G., Visu-Petra, L., Otgaar, H. (accepted). From Anecdotes To Real-Life Examples: A Mixed-Methods Survey On Self-Deception. *Cognition, Brain, Behaviour. An Interdisciplinary Journal*

consequences and impact (costs and benefits). This study took a novel approach to gathering empirical evidence on self-deception by combining qualitative and quantitative data through open-ended and multiple-choice questions. We also measured dispositional self-deception and explored its association with self-reported frequency of self-deception. We anticipated a positive correlation between dispositional self-deception and frequency of reported self-deception (reflective of mostly situational self-deception) based on previous findings that showed that situational self-deception was associated with dispositional self-deception (Chance et al., 2011). However, given the lack of previous empirical findings, this study was primarily exploratory and descriptive.

3.1.1 Method

Two hundred thirty-two students enrolled in master's and bachelor's legal psychology courses at KU Leuven (Belgium) and Babeș-Bolyai University (Romania) took part in this study as part of their practical activities in their courses. Four participants did not answer any questions; thus, only N = 228 participants were included. The age of the participants varied from 20 to 44 years old, M_{age}= 21.7, SD_{age} = 2.46. 71.9 % (n = 166) identified as women; 25.5 % (n = 59) identified as men; one participant chose the option "Other". 91.3 % (n = 211) of participants were Belgian, and 8.7% (n = 14) were Belgian–Polish, Belgian-Finnish, Dutch, German, Romanian, Brazilian, Australian, Italian and French. The number of respondents per question varied (between 107 and approximately 140 participants answered most of the questions, see Results), as the questions in the survey could be skipped if participants did not want to answer them.

Self-deception survey. A survey was constructed to explore the prevalence, examples, and characteristics of self-deception (see Appendix A and OSF for the complete survey). The survey was conducted using an online survey management database (Qualtrics, https://www.qualtrics.com), in English and contained three main parts. A *general introduction* (i.e., definition, some generic invented examples based on the literature, aim of the study; the terms "self-deception" and "lying to oneself" were used). It was emphasised that the survey referred to *participants' experiences* and not to general beliefs about self-deception. The *second part of the survey* included questions regarding socio-demographic variables (i.e., age, gender identity, nationality, and the course in which students were enrolled). The *third part of the survey* contained open-ended and multiple-choice questions regarding self-deception. These questions were constructed based on theoretical and empirical studies of self-deception (e.g., Lauria et al., 2008; Mele, 2001; von Hippel & Trivers, 2011). This part was further divided into

three sections: 1) *prevalence* questions (yes/no, frequency and recollection of emergence/onset), 2) open-ended questions focused on gathering more specific, detailed *examples* of self-deception and its elements such as theme and context, motivation, awareness and strategies of contradictory evidence, emotions associated and consequences or impact) and 3) *multiple-choice questions* regarding more general experiences of self-deception. For some of the multiple-choice questions (e.g., motivations) multiple options could be chosen and participants were asked about their intention to lie, awareness of self-deception, motivations (internal and external motivations such as feeling good, financial incentives, etc.; interpersonal vs intrapersonal), type of self-deception (denial, enhancement, etc..; situational or dispositional/trait), awareness of contradictory evidence, strategies used, emotions associated, consequences and impact (positive/negative/no effect; short-, long-term). We included multiple-choice questions to compare our results with the literature.

Dispositional self-deception. The Paulhus Deception Scales (PDS; Paulhus, 1998) were used to measure dispositional self-deception. Only the Self-Deceptive Enhancement Scale (SDE) was used. A continuous scoring total was also computed (studies comparing the two show higher Cronbach's alphas, convergent correlations with social desirability measures and better reliability, Stöber et al. 2002). The Cronbach's alpha of the SDE scale was $\alpha = .68$.

3.1.2 Results and Conclusions

As expected the prevalence of self-deception was high**.** Two hundred nineteen participants answered the prevalence question; 91.3% (n = 200) reported engaging in self-deception, while 8.7% (n = 19) reported not engaging in self-deception. Among those who self-deceived (n = 200), 50.5% (n = 101) reported that they self-deceived "sometimes", 36.5% "rarely" (n = 73), 10.5% "often" (n = 21) and 2.5% (n = 5) did not respond (see Figure 1).

Reported instances of self-deception were categorised into 9 categories based on themes identified by the first author (focal theme described based on the thematic analysis method, Kiger & Varpio, 2020; see Table 1 for some and OSF for all examples).

Table 1

Examples of Self-Deception Categorised by Theme (e.g, Context, Topic)

Category	Example	Frequency, % (*n)*
Professional (academic)	I told myself I would work very hard for school this weekend. That I was going to fail an exam. I don't need to study some extra hours because I know everything already. When you had to go to football training and told yourself you had an injury because you did not want to go to training.	51.4% (*n* =55)
Relationships	That my grandmother was still alive. Believing that your friends are your friends even though they do not act like friends. That he wasn't cheating. That I was happy in a relationship.	15.9% (*n* =17)
Not specific/Other	That I did my best even though that wasn't the case. Wanting to have a certain personal quality.	12.2% (*n* = 13)
Health (physical/ mental)	Believing I was feeling better after I had been sick.	8.4% (*n* =9)
Lie	That something was my fault, but I said it was someone else's fault.	6.5% (*n* =7)
Self-image/Self-concept	I was telling myself that I am a gentle, social, helpful, lovely, and respected guy, and if others do not understand me, it is their fault.	2.8% (*n* =3)
Sex/Sexual orientation	That I'm straight.	1.8% (*n* =2)
Feigning illness	That I was feeling sick.	0.9 % (*n* =1)

Note. *n* = 107

Examples of motivations were coded using principles of thematic analysis (Kiger & Varpio, 2020) into categorised (see Table 2 for a few examples and OSF for all) by the first author based on the pre-made list of motivation categories used in the multiple-choice question on motivations (see Table 3).

Table 2

Examples of Motivations For Engaging in Self-Deception

Motivation	Example	Frequency, % (*n)*
Avoiding negative consequences/ punishment or something unpleasant	I lied to myself to not feel guilty about taking a break. (Doing) work for school. Because I did something bad for my parents not to be mad. So that I would not feel guilty for going to accuse someone else because of the fear of punishment. Changing the opinions of others. To not be disappointed.	27.1% (*n* = 29)
Feeling better	To calm my stress. To give myself some peace. Probably nothing serious, maybe just to make me feel a bit better at the time. To make it easier to live with what had happened.	24.3% *(n* = 26)
Self-image/	Because I do not want to be a jealous person.	17.8% (*n* = 19)

Self-concept	I wanted to be brave.	
Confidence	Maybe I would believe it and do it.	15.9% (*n* = 17)
	To feel more confident about my intellectual capacities.	
	I lied in the first place because I didn't want other people to know my grades.	
	Get through the event.	
Gaining benefits	To win the argument.	8.4% (*n* = 9)
	Presents.	
Not specific/Other	Love.	6.5% (*n* = 7)
	[denial] Not wanting to see the reality.	

Note. n = 107.

Table 3

Self-reported Motivations for Engaging in Self-Deception

Motivation	Frequency, % (n)
To feel better emotionally	21.2% (*n* = 79)
To feel more confident	18.2% (*n* = 68)
To protect oneself from the truth	16.6% (*n* = 62)
Cognitive discrepancy/dissonance	8.9% (*n* = 33)
To receive approval and appreciation/validation from others	8.9% (*n* = 33)
Lack of resources (e.g., emotional)	7.2% (*n* = 27)
To avoid punishment/a negative consequence	6.7% (*n* = 25)
To gain/receive something in exchange	3.5% (*n* = 13)
To be more convincing when lying to others	3.0% (*n* = 11)
To have to remember fewer details	2.1% (*n* = 8)
Residual effect of lying	1.2% (*n* = 7)
Maintaining moral self-concept	1.2% (*n* = 7)
To receive financial incentives	0% (*n* = 0)

Note. n = 373, multiple answers possible.

When asked if self-deception was situational (*n* = 136) or was a trait (often, more than two/three times per year) *(n* = 138), the majority of participants reported it did not characterise them as a trait ("no", 77.5%, *n* = 107; "yes", 22.5%, *n* = 31), but rather they mostly self-deceived in specific situations ("yes", 75%, *n* = 102; "no", 25%, *n* = 34).

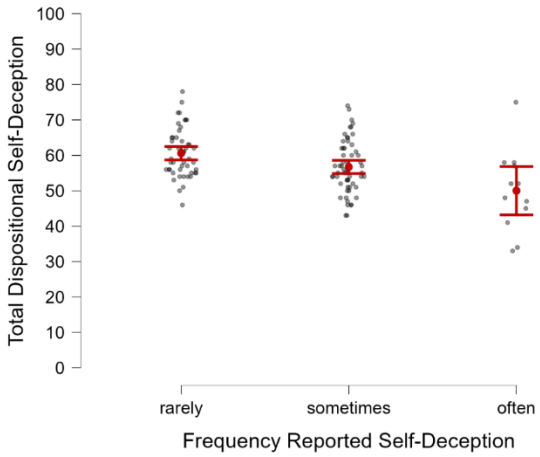
The total score for the SDE Scale (*n* = 123) was *M* = 57.74, *SD* = 8.18 (scored as continuous ranges between 20-100). A dichotomous score was also computed (*M* = 1.6, *SD* = 0.6) to compare with normative data in a larger age range sample from the general population (US/Canada, *M = 2.20, SD = 2.30, N* = 441, Paulhus, 1998). An independent-sample *t*-test was conducted to compare the means of the two groups. The present study sample and data reported on college students showed that dispositional self-deception in our study was statistically significantly lower ($t(562) = -3.57$, $p < .001$, Cohen's $d = -0.30$). The difference was statistically meaningful with a 58.4% chance that a person picked at random from the "normative" group

would have a higher score than one picked at random from the group we collected data in our study (effect size interpretation, Magnusson, 2023).

We also looked at the associations between dispositional self-deception and frequency of reported self-deception. A point-biserial correlation was conducted between dispositional self-deception (continuous and dichotomous) and reported frequency (sometimes, often, rarely), which showed a statistically significant negative correlation $rpb(121) = -.38$, $p < .001$, respectively, rpd (122) = - .21, $p < .021$ (see Figure 3). These results suggest that higher dispositional self-deception was associated with a lower reported frequency of self-deception.

Figure 3

Association between Dispositional Self-Deception and Reported Frequency of Self-Deception



Note. Error bars represent 95% CI (confidence intervals).

We explored the prevalence of self-deception, gathered examples of everyday situations of self-deception, and examined its most relevant elements and characteristics postulated and studied in the literature. For this, we created a survey with open-ended and multiple-choice questions in which participants were asked to freely write down and answer multiple-choice questions describing their experiences of self-deception. Participants reported context-specific examples of self-deception (e.g., related to their academic career and professional interests, relationships), oriented to themselves (related to their self-image), or to others (lying). Our findings supported proposed theories of self-deception, while also revealing potential areas for future research. In the following sections, we will elaborate on these findings in more detail.

One key finding of this study is the *prevalence* and *frequency* of reported self-deception. This study provides the first quantitative evidence of self-deception's prevalence, frequency, and *onset*, addressing a gap in the literature (Mele, 2001; von Hippel & Trivers, 2011). 91.3% of participants reported engaging in self-deception to varying degrees (often 10.5%, sometimes

50.5%, rarely 36.5%). Most participants (74.9%) recalled first engaging in self-deception between the ages of 6-20, with only 11.3% reporting self-deceiving before 6 years old. These findings align with developmental research on lying and false memories, showing that children can tell lies starting at two and a half years old (Białecka-Pikul et al., 2022).

Another key finding was that participants reported that they engaged in self-deception mostly in *specific situations* and were less characterised by self-deception as a *trait* or *disposition*. These findings were corroborated by the examples provided by participants, mainly referring to various contexts (e.g., professional, relationships, etc.). However, dispositional self-deception scores were statistically significantly lower than in the normative data from the general population (Paulhus, 1998), which limits our ability to generalise this finding as it could be a sample characteristic or emerge as a limitation of our method (only including participants who reported engaging in self-deception). Future studies should replicate these findings across various age groups. Nonetheless, it highlights the importance of assessing specific examples and situational self-deception, especially as many of the available studies investigating self-deception, even more in settings such as the legal field, measure mainly dispositional self-deception (e.g. for a review on sex offenders see Tan & Grace, 2008). A surprising finding was the *negative association between self-reported situational and dispositional self-deception*. Higher trait self-deceivers reported less situational self-deception, while higher situational self-deceivers reported lower trait self-deception. One explanation is that situational self-deception, viewed as undesirable, may be underreported by those with high dispositional self-deception (Paulhus & Reid, 1991). Alternatively, individuals reporting more situational self-deception may be more self-aware and honest in their evaluations, avoiding extreme scores on self-deception scales (Lu & Chang, 2011).

Despite its limitations, this study offers valuable contributions to understanding self-deception. The topic has been debated for decades, with limited progress in defining its elements and characteristics. By employing a mixed-methods approach that combines qualitative and quantitative data, this study provides insights that can help refine existing theories, address methodological debates, and support the development of new paradigms incorporating additional aspects of self-deception. Practically, the study offers an overview of self-reported real-life examples of self-deception, which could serve as a foundation for exploring its role in other contexts.

3.2 Study 2: Motivated False Beliefs? Self-concept Maintenance, Self-deception, and Memory Errors[6]

As self-relevant information is actively and systematically processed (Greve & Wentura, 2010), it could, in certain cases, lead to false beliefs (von Hippel & Trivers, 2011). A specific case of false beliefs is the phenomenon of self-deception (Balcetis, 2007; Chance et al., 2011, 2015; Mele, 2000; von Hippel & Trivers, 2011; Jian et al., 2019; Mei et al., 2023). In the present study, our main aim was to investigate internal motivations for self-deception, more specifically, self-concept maintenance. Several other internal motivations for self-deception were measured (positive and negative affect, cognitive dissonance, subjective well-being, stress, depression and anxiety, and behavioural tendencies of approach and avoidance); however, as these do not represent the main aims of this study, materials and results regarding these are included in the Supplemental Materials. A second aim was to explore the effect self-deception has on subsequent memory, in terms of remembering one's past performance, as well as in task recognition, thus testing for memory errors in self-report and actual performance.

To address these aims, a modified version of the forward-thinking paradigm was used (Chance et al., 2011), which uses a cheating context shown to elicit self-deception concerning individual performance (Chance et al., 2011, 2015; Mei et al., 2018; Ren et al., 2019). This study was done online, due to possible restrictions of the COVID-19 pandemic (data collection took place during 2021-2023). In our study, participants received a general knowledge test (TRIVIA-type questions), with half of the participants having access to the correct answers and being later asked about their estimations of their performance for a future hypothetical test (situational self-deception). We expected participants who have access to the answers to cheat and have higher situational self-deception. While cheating was not the main focus of this paradigm, the context of cheating is relevant for the legal implications that self-deception might have. Additionally, different cheating paradigms are being used in other studies that are looking at processes with implications in the legal arena, such as ethical amnesia (forgetting that they cheated, which is attributed to motivated forgetting, Shu et al., 2011; but see also Riesthuis et al., 2022 and Stanley et al., 2018) and detecting deception (e.g., Gevens et al, 2020).

One major challenge in studying self-deception is distinguishing between a genuine false belief in the actions reported and strategic dishonesty, as well as ensuring that participants are not engaging in conscious deception when reporting beliefs or behaviours. Furthermore, some

[6] The content of this subchapter is a manuscript under preparation for submission to publication. The authors are: Moldoveanu, M.G., Bulbuc, A., Otgaar, H, & Visu-Petra, L.

studies rely on indirect measures (e.g., measuring narcissism as a proxy of self-enhancement; Guenther et al, 2023), making it difficult to interpret the findings as applied to self-deception (Mijović-Prelec & Prelec, 2010). Thus, including additional measures such as attributions for task performance and self-relevance (Robins & Beer, 2001), as well as an extended debriefing, could offer a more straightforward perspective on self- vs. other-deception. Similar to previous studies (Chance et al., 2011; Chance et al., 2015; Mazar et al., 2008), we anticipate that individuals with higher self-deception will attribute their performance to personal characteristics (e.g., intellectual ability) while more honest participants will attribute their performance to having access to the answers (e.g., choose the option "other" within a list which is left intentionally open to interpretation), even if they overclaim their performance.

Additionally, we expected that higher self-deception (more positive beliefs about oneself) would be positively related to situations' self-relevance (Robins & Beer, 2001). Furthermore, at the debriefing, self-deception was further probed by asking participants if they indeed cheated and if they genuinely believed their estimations. Moreover, we measured dispositional self-deception, and we expected that situational self-deception would be associated with dispositional self-deception. We also looked at estimations immediately after the test, thus presenting estimations and exploring if they are different from the future estimations, expecting that there will be a significant difference with higher estimations for future performance.

To test the hypothesis that self-deception was associated with the maintenance of moral self-concept, the present study examined the associations between situational self-deception and moral self-concept measured one week before and immediately after the possibility of cheating (moral self-concept maintenance). Moreover, we also measured dispositional self-deception, as a previous study showed that moral self-concept was associated with dispositional self-deception (Lu & Chang, 2011). Based on the limited research, we expected that dispositional and situational self-deception were associated with moral self-concept at baseline and with moral self-concept maintenance after a cheating opportunity.

Furthermore, situational self-deception in the forward-thinking paradigm was measured as a positive future inflation. Furthermore, a well-known self-serving memory bias has been previously documented in previous research (Shacter et al., 2023; Zhang et al., 2018). We expected to find evidence for self-serving memory bias in terms of inflated beliefs in higher past performance. While initial research showed that the self-serving memory bias might not be related to present estimation or dispositional self-deception (Gramzow & Willard, 2006), different types of estimations were compared (different types of tests), which could have accounted for the results. As in our study, both the future and past estimations refer to a similar

performance context, we investigated in the present study if a higher situational self-deception was associated with an inflated past performance estimation, expecting a positive association between the two. Additionally, we also investigated if there was a positive association between dispositional self-deception and inflated past performance estimation, also referred to as self-serving memory bias, throughout this study.

Lastly, we explored whether memory errors (inconsistencies between initial answers and those given at the recognition test) were associated with situational self-deception. To our knowledge, no previous studies have investigated this association, so no predictions were made. However, based on the literature on the impact of self-generated lies on memory (Riesthuis et al., 2022, 2023), it could be the case that situational self-deception, which could be defined as a self-generated internalised lie, might be associated with memory errors.

3.2.1 Method

The procedure was approved by the ethical review board of Babeș-Bolyai University, Romania (Reference Number: 6084 / 11.06.2021) and a total of 137 participants completed all phases of the study (M_{age} = 23.72, SD_{age} = 6.43; varying between 18 and 49 years old, with 83.94% (n = 115) of the participants identified as cisgender women, 13.87% (n = 19) as cisgender men and 3 (2.19%) indicated "other" for their gender identity). The sample consisted mostly of students from various domains, such as psychology (n = 80), informatics (n = 10), economics (n = 35), and others (medicine, law, engineering, n = 3).

Participants were randomly assigned to two groups: control (n = 66; received an initial general knowledge test without answers) and cheating possibility (n = 62; received an initial general knowledge test with the correct answers), and the study was conducted entirely online. The group sizes are not equal due to participants dropping out after the second session (e.g., due to various personal reasons).

General Knowledge Tests (Test 1 and Test 2). For this task, named for participants "Trivia game", questions were selected from various general knowledge and general TRIVIA websites or games (e.g., https://www.randomtriviagenerator.com/). The test had 20 multiple-choice general knowledge questions with one correct answer per question, with a point offered for each question; thus, a maximum score of 20 could be accumulated. Participants did not receive direct feedback regarding their scores; however, the cheating possibility group could use the correct answers as feedback for their performance. Two equivalent versions of the general knowledge test were created, containing the same number of questions with a similar level of difficulty, so they can be used for each participant as Test 1 and Test 2. A pilot study

with 11 participants (N = 11 students) was conducted to evaluate these versions, ensuring they could be used interchangeably. These two versions were further counterbalanced between conditions and test orders to avoid any order or version effect.

Performance Estimations for present, **future** (self-deception), and **past** (self-serving memory bias**)** were measured. Present and past estimations focused on the first knowledge test, and participants were asked to estimate their present (immediately after the test) and past (one week later) as an overall score between 0 and 20 (as the knowledge test had 20 questions in total). This represented the total answers that the participants believed they answered correctly. Situational self-deception was operationalised as the genuine inflated false belief in a hypothetical future test, similar to the previous one, without any possibility to cheat (Chance et al., 2011). Participants were asked how they think they will perform in a future hypothetical test in which they could not cheat under any circumstances, using an overall prediction ranging between 0 and 20 (total possible score). Additionally, confidence in their estimations was measured on a Likert Scale from 1 to 10 (not confident at all to extremely confident) (See Supplementary Materials on OSF).

Moral Self-Concept And Maintenance. The Six-Factor Self-Concept Scale (SFSCS; Multidimensional Self-Esteem Scale, Stake, 1994) was used to assess the participants' moral self-concept, before (as baseline) and after a cheating scenario. For this study, the Morality subscale was used (item example: "I am a reliable person") with adequate internal consistency (Cronbach's $\alpha = 0.61$ at baseline and Cronbach's $\alpha = 0.59$ for the updated measure).

Dispositional Self-Deception. The Paulhus Deception Scales (PDS; Paulhus, 1998) were used to measure trait exaggeration response style, specifically enhancing positive qualities (dispositional self-deception). The internal consistency of the scale and subscale was adequate, with Cronbach's $\alpha = 0.56$ for the entire scale, respectively 0.78 for the SDE subscale and 0.81 for the IM subscale.

Recognition Task and Memory Errors. The recognition task consists of 20 general knowledge questions, the same questions and answers as their first knowledge test. For this, both the questions and the answers from the first test were randomised to avoid order effects, and also to ensure that the participants could not use the previous notes with the answers. For this task, it was explained to the participants that the correct answer to the recognition test was not the correct answer to the question; the correct answers were considered the answers participants gave in their previous test (e.g., correct recognition was considered choosing at Test 1 and recognition test the same answer; e.g., test 1 had the answer "B) Pablo ", correct recognition test answer "C) Pablo", incorrect was "B) Richard").

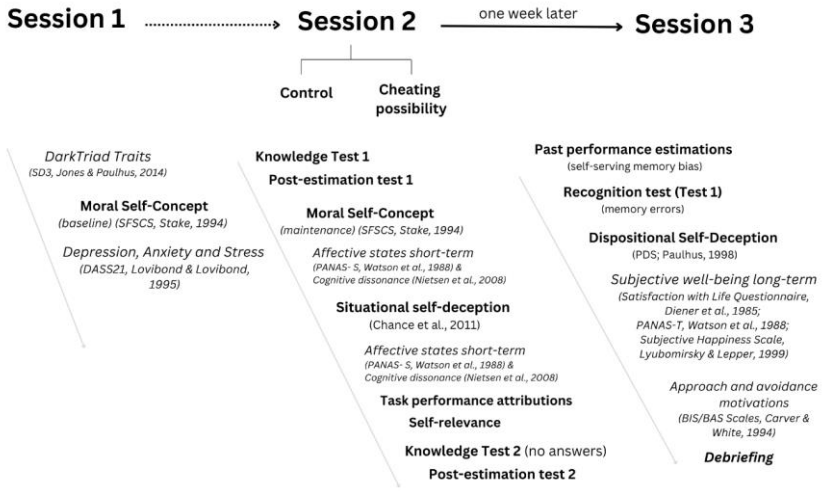
Cheating Paradigm: Design and Procedure

We used an experimental design with an experimental group (cheating possibility through receiving access to knowledge test questions) and a control group (no access to answers). Main measures (within) which taken were measures of actual performance on a knowledge Test (Test 1, with/without access to answers), present estimations, prediction of future performance (situational self-deception), actual score on Knowledge Test (Test 2, no answers given), estimation on the second knowledge test (Test 2 present estimation), as well as one week later past performance estimation (self-serving memory bias) and a recognition test score (total errors).

The study had three sessions (see Figure 2): 1) completing online demographic data and the baseline for moral self-concept, as well as a scale for Dark Traits, stress, depression and anxiety (see Supplementary Materials on OSF), 2) a one-on-one online experimental session with the researcher with the main focus on eliciting self-deception (positive self-estimations of future performance, task attributions, and self-relevance) using a cheating paradigm, moral self-concept maintenance, short-term affective states (positive, negative and cognitive dissonance, see Supplementary Materials on OSF) and 3) a second one-on-one online session that is focused on memory (past performance estimations, memory recognition test), dispositional self-deception, other possible internal motivations (subjective well-being, approach and avoidance motivation, see Supplemental Materials on OSF) and debriefing.

Figure 2

The Extended Procedure Of The Current Experiment



Note. The main measures included in this study are indicated using a bold font. These are: Moral self-concept maintenance (baseline and maintenance), knowledge Test 1 (score and estimation), situational self-deception (along with task performance attributions and self-relevance), knowledge Test 2 (score and estimation), past performance estimation, recognition test, and dispositional self-deception. Other measures presented in italics are included in the Supplemental Materials.

3.2.2 Results and Conclusions

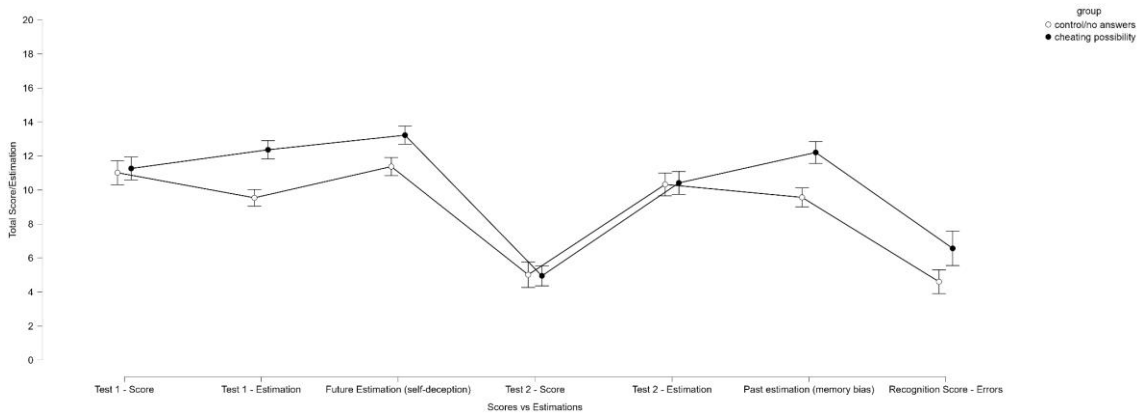
A repeated-measures ANOVA was performed to assess the effect of group (control/no answers vs cheating possibility; between-subjects factor) and type of score (actual performance score and estimations; within-subjects factor at Test 1, Future Prediction (estimation), Test 2, and Recognition Test). Mauchly's test indicated that the assumption of sphericity was violated, $\chi 2$ (20) = 228.60, $p < .001$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.63$). Levene's test indicates that group variances were statistically similar for all measures (except for Recognition score as total errors, $F(1, 119) = 6.35$, $p = 0.013$), indicating that the assumption of homogeneity of variance was met for all other measures. For the Recognition Score, an additional Welch's *t*-test was performed to corroborate the ANOVA findings.

Results of the repeated measures ANOVA test with Greenhouse-Geiser correction, indicated a statistically significant group x measure type interaction ($F(3.76,119) = 7.12$, $p < .001$, $\eta^2_p = .57$. There were also statistically significant main effect of measure type (scores vs. estimations), $F(3.76, 119) = 156.65$, $p < .001$, $\eta^2_p = .06$, and group, $F(1,119) = 22.42$, $p < .001$,

η^2_p = .16. These results indicated that the difference between measures (score/estimation) varied depending on the type of measure (if it is an actual performance score or estimation) and group (control vs cheating possibility) (see Figure 3).

Figure 3

Group Differences in Actual Scores and Estimations Across Knowledge and Recognition Measures (Test 1, Test 2, Self-deception, Recognition Test)



Note. Mean actual scores and self-estimations for Knowledge Test 1, Self-Deception Prediction, Knowledge Test 2, and Recognition Test are shown separately for the control group and the cheating possibility group. Error bars represent standard deviations.

Contrary to expectations, a simple main effect analysis of group comparing the Test 1 scores between the two groups showed that the difference was not statistically significant, $F(1, 119) = 0.32$, $p = .574$, Cohen's $d = 0.10$, thus, there was insufficient evidence to conclude that participants engaged in cheating on the first test[7].

Situational Self-Deception (Future Predictions)

To examine whether having access to answers leads to a difference in future hypothetical predictions, which is the main measure of situational self-deception, a simple main effects analysis was conducted for the prediction estimations for both groups. The difference between groups was statistically significant, $F(1, 119) = 10.97$, $p = .001$, Cohen's $d = 0.61$. Results indicated that participants with access to answers made higher future performance predictions, and thus had higher situational self-deception.

[7] Additionally, for both knowledge tests (Test 1 and 2), we also examined differences between types of questions (easy, medium, difficult and impossible) and between the two test versions to investigate further evidence of cheating as cheating; however, no statistical significant differences were found, thus to simplify the results, these are not included (see Supplemental Materials).

Situational and Dispositional Self-Deception. Situational self-deception was statistically significantly positively correlated with dispositional self-deception ($r(64) = .45$, $p < .001$) and negatively correlated with impression management ($r(64) = -.24$, $p = .047$) for the control group. However, for the cheating possibility group, situational self-deception was not statistically significantly associated with either dispositional self-deception ($r(60) = -13$, $p = .284$) or impression management ($r(60) = .14$, $p = .277$). These results indicated that the two measures of situational and dispositional self-deception were similar in the context without feedback (access to answer) and measured inflating one's self-perception in the absence of information regarding true abilities, behaviours or attitudes. However, this was not the case when more diagnostic information was available (e.g., answers pointing to one's accuracy).

Past Estimation (recognition test estimation as self-serving memory bias). The difference between groups in past estimations was statistically significant, $F(1, 119) = 19.25$, $p < .001$, Cohen's $d = 0.80$. As expected, the results indicated that a self-serving memory bias was elicited for the cheating possibility group, which estimated their past performance as higher than the control group.

Past performance estimations were statistically significantly different from situational self-deception ($t(119) = 6.57$, $p < .001$, Cohen's $d = 0.50$). These results indicated that participants' estimations of their past performance were similar to their present estimations one week later. However, participants' past performance estimations were statistically significantly different from previous elicited situational self-deception.

To test whether estimations for past estimations were associated with present and future estimations, Pearson correlations (Bonferroni-corrected significance threshold at $p < 0.00625$) were conducted, per group. The past performance estimations were statistically significant correlated positively with situational self-deception both for the control ($r(64) = .80$, $p < .001$) and the cheating possibility group ($r(60) = .65$, $p < .001$), and with the present estimations at Test 1 (control, $r(64) = .80$, $p < .001$; cheating possibility, $r(60) = .83$, $p < .001$) and Test 2 (control, $r(64) = .53$, $p < .001$; cheating possibility, $r(60) = .36$, $p = .004$). These results indicate that, independent of having or not having access to answers, participants' estimations of their past are associated with their present and future estimations, thus higher inflated past performances are associated with higher inflated present and future estimations.

Memory Errors

As expected, there was a difference between groups during recognition, with the cheating possibility group performing worse (remembering fewer of their own choices) and having more errors than the control group, as the difference was statistically significant, $F(1, 119) = 14.82$,

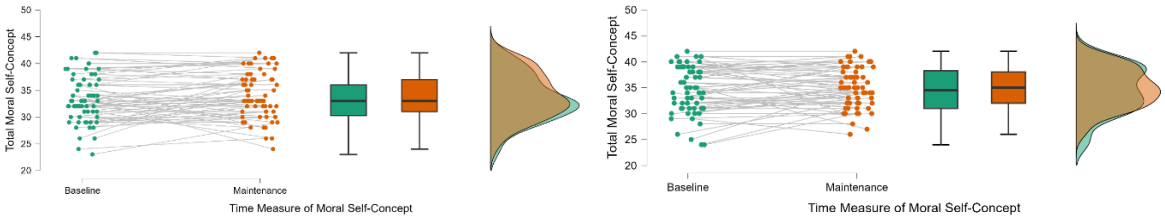
$p < .001$, Cohen's $d = 0.71$. This suggests that having access to the answers was associated with poorer recognition performance.

Memory Errors, Self-Deception, and Self-Serving Memory Bias. In the control group, the correlations between situational self-deception, self-serving memory bias and memory errors were not statistically significant ($r(65) = 0.04$, $p = .769$, respectively $r(65) = 0.00$, $p = 1.00$). In the cheating possibility group, the amount of memory errors and situational self-deception were not statistically significantly associated ($r(62) = -.01$, $p = .956$), nor were self-serving memory bias and memory errors ($r(62) = -.27$, $p = .030$). Results indicate that inflated future or past estimations are not associated with memory errors.

Moral Self-Concept Maintenance And Self-Deception. Contrary to expectations, there was no statistically significant interaction effect, moral self-concept measure type x group effect, $F_{(1,124)} = 0.296$, $p = .587$, $\eta_p^2 = .002$, and no statistically significant difference between the two measures of moral self-concept (baseline measure: control, $M = 33.29$, $SD = 2.32$; cheating possibility $M = 34.52$, $SD = 5.56$; maintenance measure: control, $M = 34.00$, $SD = 4.29$; cheating possibility, $M = 34.88$, $SD = 3.80$), $F(1,124) = 2.886$, $p = .092$, $\eta_p^2 = 0.02$), nor between the two groups, $F(1,124) = 2.34$, $p = .129$, $\eta_p^2 = 0.02$ (see Figure 5). These results suggest that moral self-concept was similarly maintained for both the control and the cheating possibility group, and having a cheating possibility did not lead to differences in moral self-concept maintenance.

Figure 5

Raincloud Plots for Moral Self-Concept Measures by Group



Note. Control group (left) and cheating possibility group (right). Moral self-concept measures: baseline, approximately one week before the knowledge test with/without cheating possibility, and maintenance immediately after the knowledge test with/without cheating possibility.

Moral self-concept at baseline for the control condition did not correlate significantly with situational self-deception, $r(64) = .29$, $p = .019$, but correlated significantly positively with dispositional self-deception, $r(64) = .33$, $p = .007$. Moral self-concept after the first knowledge test (maintenance) only correlated with trait dispositional self-deception for both groups (control, $r(64) = .36$, $p = .003$ and cheating possibility, $r(61) = .33$, $p = .009$). There was no significant correlation between moral self-concept maintenance and situational self-deception

for either group. For the possibility to cheat group, moral self-concept (before and after the cheating possibility) was associated with dispositional self-deception ($r(60)=.26$, $p = .045$). These results suggest that while individuals with higher dispositional self-deception reported consistently higher moral self-concept (both at baseline and after the cheating possibility task). Situational self-deception was not associated with the baseline measure of moral self-concept. Moral self-concept maintenance was not associated with situational self-deception (with/without cheating possibility).

Tabel 3

Descriptives and Correlations Between Self-Deception (Situational and Dispositional) and Moral Self-Concept By Group

	Control (*n* = 66) *M (SD)*	*Cheating possibility* (*n* = 62) *M (SD)*	1	2	3	4
1. Situational self-deception	11.32 (3.22)	13.20 (2.99)		-.14	.20	.24
2. Dispositional self-deception	61.45 (10.14)	62.98 (9.14)	.45**		.26*	.33**
3. Moral self-concept (baseline)	33.29 (4.32)	34.57 (4.54)	.29	.33**		.63**
4. Moral self-concept (maintenance)	34.00 (4.29)	34.80 (3.82)	.23	.36**	.68**	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Dispositional self-deception was calculated using a continuous approach. The control group is presented below, and the cheating possibility group is presented above.

Our primary objective was to examine internal motivations underlying self-deception, with a particular focus on maintaining a positive moral self-concept. Additionally, we sought to determine how self-deception influenced subsequent memory, specifically regarding the recall of one's own performance (self-serving memory bias) and the ability to recognise previously made choices. To address these questions, we employed a modified version of the forward-thinking paradigm (Chance et al., 2011). Unlike the original in-person design, our study was conducted online. Participants were randomly assigned to either receive access to a knowledge test or not. Measures of moral self-concept were collected both before and after the test. We assessed both situational and dispositional self-deception and included tasks to evaluate participants' estimations of their past performance as well as their recognition of test items, to explore the potential impact of self-deception on memory. The present study found a significant interaction between group (control vs. cheating possibility) and measure type (actual scores vs. self-estimations), indicating that the relationship between participants' actual performance and

their self-estimations depended on whether they had the opportunity to access correct answers. The significant interaction between group and measure type found in this study demonstrates that the effect of access to correct answers on self-estimation is not uniform across all participants or measures. Specifically, participants who had the opportunity to check for the correct answers significantly overestimated their performance compared to participants who did not have access to answers, and this effect was evident primarily in their self-estimations, not their actual scores.

First of all, in our study, we did not find evidence of cheating in terms of higher scores for participants who had access to answers, contrary to other studies (Chance et al., 2011, 2015; Fan et al., 2022; Jian et al., 2019; Ren et al., 2018; Mei et al., 2023). As no cheating was proven, we will mostly discuss our findings in terms of having or not having access to answers. Even in the absence of evidence of overt cheating, we did, however, successfully elicit self-deception. Participants who had extra access to answers showed more positive, inflated false beliefs regarding their future performance, while performing similarly in the first and the second knowledge test, a finding consistent with previous studies (e.g., Chance et al., 2011, 2015; Jian et al., 2019; Liu et al., 2019, 2025; Mei et al., 2023). As in previous studies (e.g., Chance et al., 2011, 2015), we examined the association between situational and dispositional self-deception. Consistent with previous studies (e.g., Chance et al., 2011, 2015), there were positive associations. However, the positive association in our study was significant only for the group that did not have access to answers. Overall, our results suggest that situational and dispositional self-deception are linked, particularly when measuring inflated estimates of future and general abilities in contexts without performance feedback or possible self-affirming contexts.

Our second most important finding is related to the maintenance of self-concept. In this study, we found a stable, maintained, moral self-concept consistent with a prior study which used similar self-deception (Mazar et al., 2008). Changes in moral self-concept did not statistically differ between groups and also nor between the two measures. Research on moral self-image shows that while the moral self-concept is relatively stable, some situational factors might influence it (Jordan et al., 2015). Most people declare themselves to be mostly moral individuals (Aquino & Reed, 2002). Furthermore, as in a previous study (Lu & Chang, 2011), moral self-concept was associated with dispositional self-deception, independent of group, thus, individuals with a higher tendency to inflate positive qualities about self and deny negative qualities were also presenting themselves as more moral. In terms of situational self-deception, the association with moral self-concept was not significant for either group. These findings suggest that initial, stable perceptions of one's moral self may be more closely linked to

dispositional self-deceptive tendencies. Lastly, while the positive association between dispositional self-deception and moral self-concept could represent actual associations, it is important to be further investigated, as one important critique might be that the two measures have a confounded effect of positivity (Wojcik & Ditto, 2014).

Lastly, our third main finding refers to our aim to explore the effect that eliciting self-deception in a cheating context had on memory by looking at self-serving memory bias and memory errors in a recognition task in which participants either remembered correctly, forgot or chose a different answer than one week before. We found a self-serving memory bias as some participants remembered their performance more positively, which is consistent with the literature (D'Argembeau & Van der Linden, 2008; Ritchie et al., 2017; Schacter et al., 2023; Sedikides et al., 2016; Sedikides & Green, 2000; Zhang et al., 2024). When we examined memory errors, having access to answers led to more errors in remembering previous choices. However, these errors were not associated with inflated estimations in terms of situational self-deception, self-serving memory bias, or cheating, as in lying about choices.

This study showed that when offered an opportunity to cheat (even without actual cheating), participants engaged in successful situational self-deception and self-serving memory bias, thus predicting and recalling their performance more favourably than it was. Moreover, while dispositional self-deception and moral self-concept were positively associated, moral self-concept maintenance was not associated with situational self-deception. As in a previous study (Lu & Chang, 2011), dispositional self-deception and moral self-concept were positively associated. Memory errors increased when offered a cheating opportunity (access to truthful information), but were not associated with self-deception as measured in the present study. Future studies could use methods that include the ground truth and measure self-deception in the context of past events or details besides specific evaluations of one's abilities. These findings highlight the importance of understanding self-deception and memory bias, especially in contexts like legal psychology, where motivated false beliefs can have significant real-world effects.

3.3 Study 3: The Effect of Plausibility and Event Frequency on the Implantation of False Beliefs and Memories [8]

Over the past decades, it has become widely accepted that memory, while highly accurate in general, is reconstructive and can be prone to distortion (Loftus, 2005; Otgaar et al., 2022b). For example, research has shown that people can form false memories (memories of non-experienced events/details) of autobiographical experiences (Garry et al., 1996; Mazzoni & Memon, 2003; Otgaar et al., 2012). A crucial distinction here is between autobiographical memory (recollection) and autobiographical belief (belief in an event's occurrence, Scoboria et al., 2004; Smeets et al., 2005). To investigate suggestion-based false memories, researchers have, for example, used the false memory implantation paradigm (Loftus & Pickrell, 1995). In this paradigm, based on participants' parents' reports of experienced childhood events, participants receive the suggestion that they experienced a false event. The typical finding is that a proportion of participants (about 30%; e.g. Scoboria et al., 2017; see also Wade et al., 2025 for a critical discussion) form false memories. However, this paradigm has several limitations: training for conducting suggestive interviews, differences in false memory rates due to scoring methods (e.g., Calado et al., 2021; Wade et al., 2018), and the need to rely on others for establishing event occurrences.

Otgaar et al. (2022b) introduced and validated a novel method, namely the blind implantation technique. This method was inspired by the literature on memory and choice blindness (people not noticing when alterations are made in their memory reports, Cochran et al., 2016; and in their self-made choices; Sagana et al., 2014). In this paradigm, participants receive a list of autobiographical events (20 events, one critical) and are asked to indicate whether they have experienced them. One week later, participants who indicated they had not experienced the critical event received a shorter, personalised list of events. They are told that these are the events they stated as having been experienced (four true and one false critical event). Participants are then asked to provide belief and recollection ratings and to recall additional details of these events. Otgaar et al. (2022b) implanted false beliefs and false memories ranging between 9% and 30%.

Previous studies examining the effect of plausibility on false memories showed that children were more likely to create false memories for plausible events (e.g., getting lost in a shopping mall) than implausible ones (e.g., receiving a rectal enema) (Pezdek & Hodge, 1999).

[8] The content of this subchapter is a manuscript under revision, prepared for publication. The authors are: Moldoveanu, M.G., Shahvaroughi, A., Manguilli, I., Hatami, J. & Otgaar, H.

Pezdek and Hodge (1999) presented children with two false events (plausible and implausible) and later interviewed them regarding their memories of these events. Out of 39 participants, 35.9% (n = 14) recalled the plausible event, 7.7% (n = 3) recalled both false events and one participant (2.6%) recalled the implausible event. Conversely, some studies investigating plausibility and false memories in children found that event plausibility did not significantly affect children's false memory formation (Strange et al., 2006; Otgaar et al., 2009). These inconsistent findings may be due to varying methodologies (different scoring methods, use of critical events). Furthermore, several studies with adults showed that implausible events become more believable with increased exposure (e.g., imagery techniques, Bays et al., 2012; visualisations, Blandón-Gitlin & Gerkens, 2010 and information, Mazzoni et al., 2001), which could in turn lead to false memories (Sharman & Scoboria, 2008). For example, Mazzoni et al. (2001) gave participants a plausibility-enhancing manipulation (mini-articles containing accounts of the occurrence of events in the study) along with a personalised suggestion (false feedback that the events could have happened to them in the past), which led to an increase in false beliefs in the occurrence of these events. Given that abuse cases often involve multiple occurrences, a question that arises is whether less plausible repeated events differ from more plausible repeated events in their likelihood to create false memories?

The present experiment aimed to investigate the effect of event plausibility and frequency on the formation of false beliefs and false memories. We directly compared outcomes for high and low plausible events (e.g., Strange et al., 2006), focusing on the ease of implanting false beliefs and memories for high and low plausible events, as well as single and repeated events. Based on findings showing that high plausibility facilitates false memories (Pezdek & Hodge, 1999), that single and repeated events lead to false beliefs and memories (with false beliefs easier to elicit for single events, Otgaar et al., 2022b), and that personalized suggestions (false feedback about the event's past occurrence) can increase false beliefs (Mazzoni et al., 2001), we hypothesized that: (1) a low plausible event would elicit lower false beliefs and memories than a high plausible event; (2) an event that allegedly happened once would generate lower false beliefs and memories than a repeated event and (3) a repeated low plausible event would elicit lower false beliefs and memories than a high plausible happened once event.

3.3.1 Method

Data was collected from the general population (Western European residents in the Netherlands and Germany, Romania, and Iran) and students (Maastricht University, the Netherlands; Babeș-Bolyai University, Romania and University of Tehran, Iran). The first part

of the experiment (i.e., an online survey) was partially (n = 50) or fully (n = 805) completed by a total of N = 855 participants. Out of these, only participants who would not report experiencing two critical events during the first part of the experiment were eligible to continue. Thus, n = 222 (27.6%) did not experience the high plausible critical event, and a further n = 193 (22.6%) did not experience the low plausible critical event. Per the exclusion criteria, these latter participants received the second part of the experiment (i.e., if they also had experienced other 4 true events besides the critical events, they received a second personalised survey). Next, we examined the number of participants who fully completed the second part of the experiment. Our final sample consisted of N = 103 participants (data on demographics only for n = 85 was available; M_{age} = 33.70, SD_{age} = 12.26, between 18 and 67 years old; with 64 (62.1%) participants identifying themselves as women, 19 (18.4%) as men and for the rest their gender identity was not specified (n = 2; 1.9%)). These participants were used for the analyses reported below (Western Europe: n = 44; Romania: n = 9; Iran: n =50)[9]. A pilot study was conducted to identify high and low plausible events that could be used as critical events in our study. The highest-rated plausible event was "lost a toy as a child" (M = 6.22, SD = 1.38), which was chosen as the critical plausible event for this experiment. The event "almost drowning in the ocean" (M = 3.78, SD = 2.13), was selected as the low implausible event for this experiment.

Survey 1. The first part of this experiment consisted of an online survey, which contained two parts: 1) the informed consent (e.g., research cover story, confidentiality, participation benefits, etc.) and 2) a list of childhood events, which measured the initial belief in the occurrence of some childhood events. The list consisted of 20 events (e.g., "Build a snowman with a friend", see OSF and Appendix). For each event, participants had to indicate whether they experienced it in their childhood ("yes"/"no"). For the experienced events (events indicated as "yes"), another question was asked about the frequency of this event in participants' childhood (i.e., "one time", "two times", "three or more times"). For the non-experienced events (answered "no"), no additional question was asked. Of the twenty events, two events were considered critical (intended to be used as false events), with one event corresponding to the " high plausible" (i.e., "lost a toy in childhood") and one to the "low plausible" (i.e.," almost drowned in the ocean") condition. Participants had to indicate they had not experienced these

[9] We investigated whether the Western European, Romanian, and Iranian participants (categories made based on the language and location in which they received the survey and not nationality, as this was not collected) statistically differed on belief and recollection ratings for the critical false event. There was no significant difference between the 3 groups for any of the beliefs (before imagination: $F(2, 100) = 1.90$, $p = .1555$, $\eta_p^2 = 0.04$; after imagination: $F(2, 100) = 2.65$, $p = .075$, $\eta_p^2 = 0.05$ and recollection (before imagination: $F(2, 100) = 1.82$, $p = .168$, $\eta_p^2 = 0.04$; after imagination: $F(2, 100) = 1.82$, $p = .168$, $\eta_p^2 = 0.04$) ratings before and after imagination.

critical events, thus answering “yes/no” with no indication regarding the event frequency of these events.

Survey 2. The second part of the experiment consisted of a second, personalised survey that measured the belief in occurrence and the recollection of four true and one false events, before and after an imagination instruction, thus true and false beliefs and memories. It consisted of five events (“*Previously you have indicated that you experienced the event… once/more than once*”) for which participants were asked to provide their belief in the occurrence of the event during their childhood and the recollection ratings. Both belief and recollection were rated on 8-point Likert scales with belief in occurrence ranging from 1 = definitely did not happen to 8 = definitely did happen and recollection from 1 = no memory for the event at all to 8 = clear and complete memory of the event (see also Scoboria et al., 2004; Otgaar et al., 2022b). For each event, participants were given an imagination instruction, followed by a free recall task ("*Please try to imagine the event, how it happened, and try to remember as many details as possible. For example, think about how old you were, who you were with, and where the event took place. Please give your report below.*"). Participants were asked for belief and recollection ratings a second time after the imagination and free recall task for each event. The order of events was consistent for all participants (the first two events presented were true events, followed by the false event, and the last two were true events). However, the true events were different for each participant, as these events consisted of events for which the participants answered “yes” in the first survey. Lastly, the survey contained information regarding the false event (“one of the below events was not truly experienced by you”), they were asked if they recall which one was not truly experienced, followed by an 8-point Likert scale confidence rating in their opinion (1 = not at all confident to 8 = extremely confident). This survey also included a debriefing section for participants with an open-ended question regarding their idea of the purpose of the study and a text describing the actual purpose of the study (see Materials on OSF). Both surveys were created and distributed using online survey platforms (e.g., Qualtrics, https://www.qualtrics.com/).

We used a between-subjects design with Plausibility (High (n = 51) versus Low (n = 52)) and Frequency (Single (n = 55) versus Repeated (n = 48)) as the between-subjects variables. Before researchers sent the second survey, they checked whether participants met the inclusion criteria (i.e., had not experienced the false event and had experienced four true events out of 18). If participants met these criteria, researchers randomly assigned them to one of four groups (High Plausible Single (n = 25), High Plausible Repeated (n = 26), Low Plausible Single

(n = 30), and Low Plausible Repeated (n = 22)). As some participants did not return the second questionnaire, the groups used for analysis did not have equal sample sizes.

Scoring. In the current experiment, we employed the same scoring strategy as in Otgaar et al. (2022b; see also Scoboria et al., 2004, 2014; Shavahroughi et al., 2025). Specifically, both false beliefs and false memories were scored based on self-reported belief in occurrence and recollection on an 8-point Likert scale. Thus, a false belief was scored when participants had a "belief score of 5 or higher for the critical event and a recollection score lower than the belief score". Similarly, when participants had a "recollection of 5 or higher and a belief score that was the same as the recollection score"[10] it was scored as a false memory. When a false belief or false memory was also contained in the free recall, additional details, with at least one detail (consisting of specific details such as place, time, persons involved, etc.), a detailed false report was scored. For the second survey, we also looked at true beliefs and memories percentages, which were scored using the same criteria as false beliefs and false memories.

3.3.2 Results and Conclusions

Planned Analyses

False Beliefs and False Memories for High/Low Plausible and Single/Repeated Events

Table 1.

Mean and Standard Deviation (in parentheses) of False Belief and Recollection Ratings as a Function of Condition

	High Plausible		Low Plausible	
	Single	Repeated	Single	Repeated
	(*n* = 25)	(*n* =26)	(*n* = 30)	(*n* =22)
False belief before imagination	5.36 (2.33)	4.27(2.59)	2.87 (2.68)	4.14 (2.90)
False recollection before imagination	3.36 (2.41)	2.69 (1.98)	2.50 (2.22)	3.50 (2.72)
False belief after imagination	5.56(2.31)	3.73 (2.81)	2.83 (2.61)	4.27 (2.96)
False recollection after imagination	3.40 (2.42)	2.58 (2.14)	2.53 (2.33)	3.73 (2.71)

Note. $N = 103$.

False Belief Ratings Before Imagination. A statistically significant event Plausibility x Frequency interaction, $F(1, 99) = 5.15$, $p = .025$, $\eta_p^2 = 0.05$. Furthermore, our analysis revealed a statistically significant main effect of event Plausibility, $F(1, 99) = 6.38$, $p = .013$, $\eta_p^2 = .06$. No statistically significant effect was found for event Frequency ($F(1, 99) = 0.30$, p

[10] While outside the scope of our scoring, we also scored a recollection of 5 or higher and a belief lower than the recollection (i.e., nonbelieved remembered events), to explore if there are any differences in different memory types (Scoboria et al., 2004). However, as no statistical differences were found it was not included in the main results.

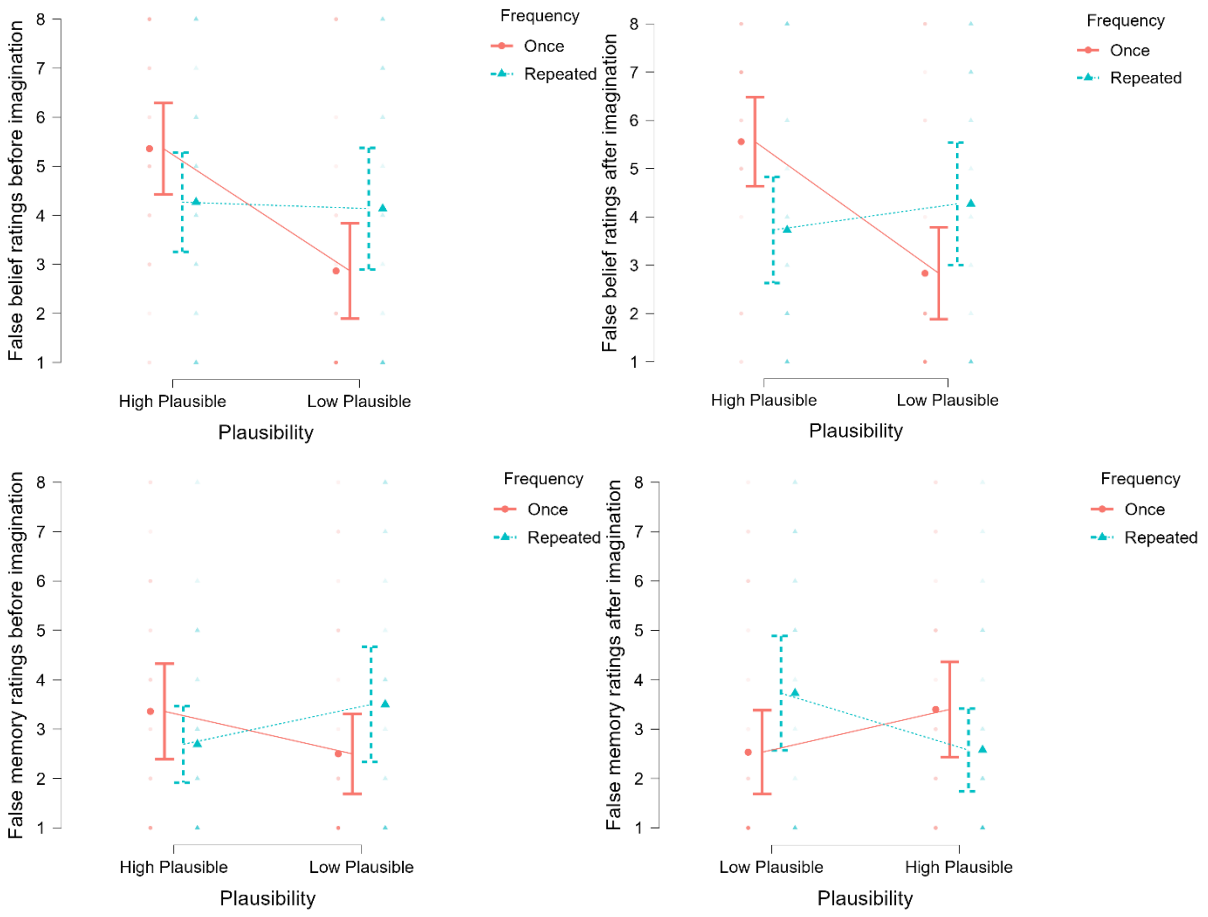
= .585, η_p^2 < .001). Simple main effects analysis showed that the effect of event Plausibility was statistically significant in the experienced Single condition (High Plausible event: *M* = 5.36, *SD* = 2.33; Low Plausible event: *M* = 2.87, *SD* = 2.68; *F(1)* = 12.36, *p* < .001, Cohen's *d* = 0.99), but not statistically significant in the Repeated condition (High Plausible event: *M* = 4.27, *SD* = 2.59, Low Plausible event *M* = 4.14, *SD* = 2.90; *F*(1)= 0.03, *p* = .862, Cohen's *d* = 0.05). Specifically, only for an event allegedly experienced once, the high plausible event led to higher false belief ratings than the low plausibility event. This effect was not present for the repeated event condition.

False Belief Ratings After Imagination. A two-way ANOVA again revealed a significant event Plausibility x Frequency interaction (*F* (1, 99) = 9.50, *p* = .003, η_p^2 = .09).. The main effect of event Plausibility remained significant but lower than before the imagination instruction (*F*(1, 99) = 4.25, *p* = .042, η_p^2 = .04), while event Frequency continued to show no significant effect (*F*(1,99) = 0.14, *p* = .714, η_p^2 = 0.001). Simple main effects analysis demonstrated that event Plausibility's impact was still significant only in the Single condition, with an even stronger effect (High Plausible event: *M* = 5.56, *SD* = 2.31; Low Plausible event: *M* = 2.83, *SD* = 2.61; *F*(1) = 14.18, *p* < .001, Cohen's *d* = 1.11). In the Repeated condition, the non-significant effect changed direction (High Plausible event: *M* = 3.73, *SD* = 2.81, Low Plausible event: *M* = 4.27, *SD* = 2.96; *F*(1) = 0.49, *p* = .486, Cohen's *d* = - 0.19). These results indicate that the imagination procedure maintained the pattern of plausibility's influence on false belief ratings, primarily affecting events experienced once, while potentially slightly altering the dynamics in the repeated event condition. Similar to false beliefs before the imagination procedure, non-parametric tests, Kruskal-Wallis, for both Plausibility (*H*(1) = 5.07, *p* = .024) and Frequency (*H* (1) = 0.02, *p* = .887) corroborate the ANOVA findings.

False Recollection Ratings Before and After Imagination. Two-way ANOVAs revealed no statistically significant event Plausibility x Frequency interaction (*F*(1,99) = 3.27, *p* = .074, η_p^2 = 0.03), nor any main effect of event Plausibility (*F*(1,99) = 0.00, *p* = .955, η_p^2 < .001) or event Frequency (*F*(1,99) = 0.13, *p* = .720, η_p^2 = 0.001) for mean recollection rating before imagination. However, there was a statistical significant event Plausibility x Frequency interaction effect for false recollection ratings after imagination (*F*(1, 99) = 4.52, *p* = .036, η_p^2 = .04), but no statistical significant effect for either event Plausibility (*F*(1,99) = 0.09, *p* = .765, η_p^2 <.001), or event Frequency (*F*(1,99) = 0.15, *p* = .697, η_p^2 = 0.001). Thus, for events experienced once or multiple times, there was no statistically significant difference in false memory ratings for both the high and low plausible event, before or after the imagination instruction.

Figure 1

Belief and False Recollection Ratings (before and after the imagination instruction) by Plausibility and Frequency of False Event



Note. Error bars in the figure represent standard errors.

Manipulation Check and Identification of the False Event

The false event identification did not statistically differ between the High Plausible, Low Plausible, Single and Repeated event groups ($\chi^2(3, 59) = 3.72$, $p = .294$, Cramer's $V = 0.25$). Furthermore, the confidence in the false event identification did not statistically differ for High Plausible, Low Plausible, Single and Repeated event groups *((*$\chi^2(9, 59) = 7.91$, $p = .543$, Cramer's $V = 0.51$).

Exploratory Analyses

Percentages of False Beliefs and False Memories

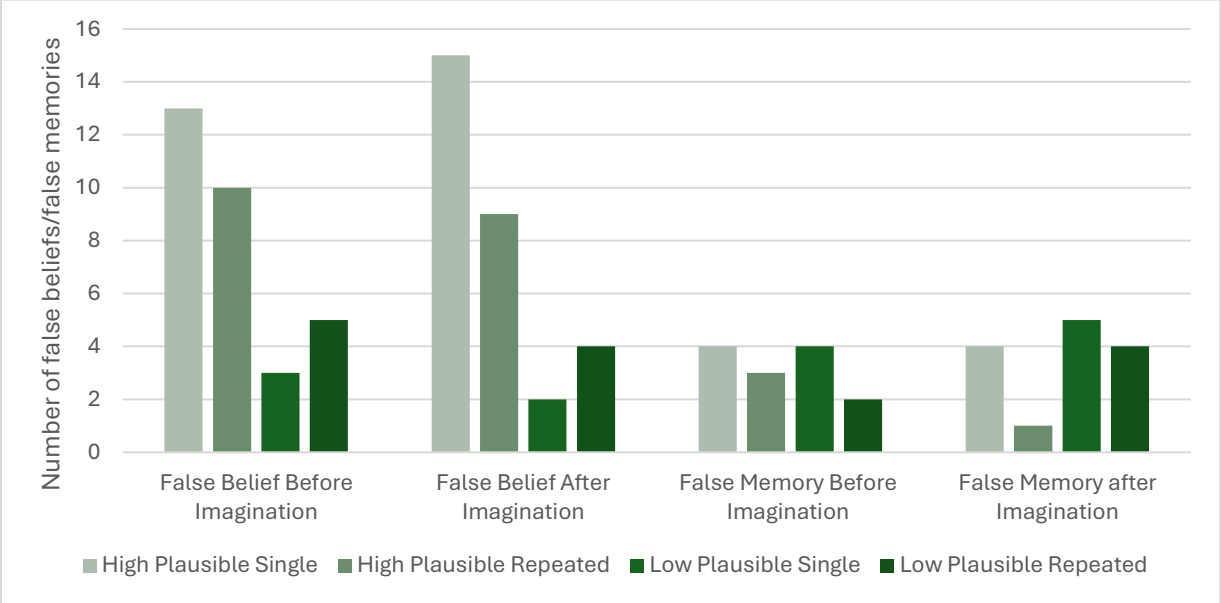
We examined the overall percentages of false belief and false memory implantation before and after the imagination instruction to replicate previous work on blind implantation (see Otgaar et al., 2022b). Our findings showed false belief and false memory rates varying

from 12.6% (n = 13) to 30.1% (n = 31) irrespective of conditions (see Table 2). Additionally, we also looked at the percentages of true event beliefs and true memories before and following the imagination instruction. These true event belief and memory rates ranged from 18.4% (n = 19) to 68% (n = 70), consistently surpassing the percentages observed for false beliefs and false memories.

False Beliefs and False Memories for High/Low Plausible and Single/Repeated Events per Group

Figure 2

Number of False Beliefs and False Memories per Participant as a Function of Condition



Note. High Plausible Single n = 25; High Plausible Repeated n = 26; Low Plausible Single n =30; Low Plausible Repeated n =22.

When looking at false beliefs before the imagination instruction, we found a statistically significant effect between groups ($\chi^2(3, 103) = 12.89$, $p = .005$, Cramer's $V = 0.35$). This result suggests that false beliefs were more easily formed in the High Plausible Single than in the High Plausible Repeated and the Low Plausible Single and Low Plausible Repeated event groups (see Table 3). Moreover, when looking at differences between groups in false belief formation after the imagination instruction, the effect was still statistically significant and even stronger ($\chi^2(3, 103) = 20.53$, $p < .001$, Cramer's $V = 0.45$).

Besides the self-reported belief and recollection ratings, we asked participants to provide written details about the events (both true and false events). For the false event, some examples of the details provided by some participants are: "*I was at the beginning of school*

when I noticed that I couldn't find a toy at my grandparents that I knew for sure I had left there. I suspected that my grandmother had either thrown it away or given it to my cousin, who had damaged it. He didn't admit anything, and we never found out. At the same time, I remember bits of my search in my grandmother's closets and other corners of the house in search of the doll." and *"I had a little teddy bear that I couldn't find at one point. But I don't remember any other.*

False memories and beliefs for suggested events can have severe real-world consequences (Loftus, 2005; Otgaar et al., 2022a). While previous research has explored factors such as event plausibility (Bays et al., 2012; Mazzoni et al., 2001; Pezdek & Hodge, 1999; Pezdek et al., 1997; Scoboria et al., 2006, 2012; Sharman & Scoboria, 2009; Strange et al., 2006) and event frequency (Calado et al., 2021; Otgaar et al., 2022b), these factors have often been studied separately. Additionally, many studies focused only on false memories without looking at false beliefs, which can lead to changes in behaviour even in the absence of recollection (Scoboria et al., 2004; Wang et al., 2019). This study examined how event plausibility and frequency affected false autobiographical belief and memory formation, using a false memory implantation paradigm, blind implantation (Otgaar et al., 2022b). Results revealed two key findings. First, plausibility and event frequency interacted in shaping false beliefs, with the single high plausible event leading to higher false belief ratings than the low plausible event. This pattern was absent for events that allegedly happened repeatedly. Second, false memory formation or recollection ratings were unaffected by our variables. We will now elaborate on our findings.

Belief and memory are distinct yet related constructs (Mazzoni & Kirsch, 2002; Scoboria et al., 2004), with important implications in theoretical and applied settings, particularly legal contexts. Our findings show that high and low plausible events can lead to false beliefs, with single high plausibility events being more likely to elicit false beliefs. Moreover, false memories can form independent of the event's plausibility and frequency characteristics. This finding underscores the need for caution when relying on reports of past experiences, especially in cases involving suggestive questioning or therapeutic techniques. Event plausibility and frequency can influence the reliability of victim testimonies, with effects different for single versus repeated occurrences. Legal practitioners should consider these factors and the possibility of false beliefs and memories when evaluating testimonies.

3.4 **Study 4: Individual Differences in Aggression and Attitudes Towards Lying** while Accounting for Social Desirability in Violent Offenders and the General Population[11]

A relatively small subset of individuals is responsible for a disproportionate amount of crime and violence (e.g., violent offenders, Falk et al., 2014; Seidl et al., 2020). **S**evere acts of *aggression* and *violence*, such as assault or murder, inflict intense harm, impacting not only the aggressor and the victim but also their communities (Parker & Morgan, 2020). Aggression, defined as behaviour intended to harm others, is a multifaceted concept explored through various frameworks **(**Anderson & Bushman, 2002**;** Komasi et al., 2016; Parker & Morgan, 2020). It ranges from momentary behaviours to stable patterns of thinking, acting, and feeling (trait aggression), along a continuum of individual differences and can differ from severe violent behaviours (Komasi et al., 2016)**.** Likewise, *lying*, defined as deliberate deception in communication, is widespread, with studies suggesting it occurs on average daily (Vrij, 2008), displaying a within-individual day-to-day variation known as "good/bad lie days" (Serota et al., 2021). While most people are generally honest (90% of all lies are small white lies; Serota et al., 2021), a small subgroup, known as prolific liars, is responsible for most lies (Serota & Levine, 2015; Serota et al., 2021, 2024). Lying and deception can harm mental health and societal trust (Vrij et al., 2008; but see also the literature on prosocial lies; Levine & Lupoli, 2022).

Aggression, violent behaviour, and deception are widely seen as undesirable traits and behaviours by society and are often categorised as signs of maladjustment**.** The individual evaluation of the desirability of a behaviour or an expressed attitude is essential, especially when looking at self-reported measurements. Research consistently shows that individuals tend to deny behaviours considered undesirable and claim socially desirable ones, a phenomenon known as *social desirability* (Crowne & Marlowe, 1960; Paulhus, 1991; Vigil-Colet et al., 2012). When measuring social desirability using self-reported scales, there is consensus that our tendencies to appear desirable can threaten the validity of these measures. However, the issue of how to handle responses during analysis is the topic of an ongoing debate. For example, social desirability can be seen as a response style, a personality trait, or a disposition (Barry et al., 2016). Most research to date treats social desirability as a confounding variable (as having style), controlling it statistically during analyses. However, during the last decade, there have

[11] [11] The content of this subchapter is a manuscript under revision, prepared for publication. The authors are: Moldoveanu, M.G., Turi, A., Hurezan, L., & Visu-Petra, L..

been strong arguments for treating this construct as a consistent trait or disposition (as having substance). The former approach could lead to a loss in meaningful variance across measures (see for a further discussion Mills & Kroner, 2005; Vigil-Colet et al., 2012). Studies that treated social desirability as a trait showed that it adds unique variance to violent recidivism (Mills & Kroner, 2005, 2006). This approach will also be adopted in the current study to investigate if it adds unique variance to aggression, violent behaviour, and attitudes towards lying.

While interpersonal violence and deception represent important issues for society, they are of even greater relevance in legal and forensic settings (e.g., correctional settings, and court processes) where, for example, aggression has been shown as being more prevalent and severe compared to the general population **(**Hornsveld & Kraaimaat, 2023; Ohlsson & Ireland, 2011). Aggression is a key factor in risk assessment measures used to establish sentencing based on the risk of recidivism (see also Matlasz et al., 2020)**,** and is central to prevention and intervention programmes aimed at decreasing undesirable and costly behaviours (Komasi et al., 2016). Equally, deception and attitudes towards lying can vary across different populations, with research to date focused mostly on lie detection abilities or beliefs about deception (Granhag et al., 2004; Schindler et al., 2020). A recent study by Hurezean and colleagues (2024) looked at Dark Traits as risk factors for reoffending and recidivism and found that deceitfulness, which represents "an inclination towards adopting deceptive, unlawful, and illicit behaviours, and disobeying the norms of civic cooperation in the pursuit of personal goals", can be a marker for first-time offending and also recidivism (Hurezan et al., 2024), making both attitudes and inclination towards deception essential risk factors for prevention and intervention programs. Nonetheless, most research on the personal factors behind aggression, violent behaviours, and deception has primarily focused on the general population, with less attention given to legal or forensic samples (such as violent offenders). Moreover, the limited research on legal or forensic populations tends to emphasise clinical and situational factors, focusing less on psychological and individual ones (Steinert, 2002).

Furthermore**, f**ew studies have examined individual differences while considering multiple predictors simultaneously, with most research focusing on one or two predictors at a time (see also Komasi et al., 2016). For offenders, particularly violent offenders, understanding the interplay between these traits and behaviours is crucial for designing targeted interventions that effectively reduce aggression and deception within correctional facilities. This would also impact the reintegration process, where triggers for these behaviours may be more prevalent, thus reducing the risk of (violent) recidivism. Similarly, in the general population, awareness of these factors is vital for developing programs that promote a sustained decrease in aggressive

and deceptive behaviours in everyday settings, where such undesirable behaviours may be normalised or reinforced (Komasi, 2016; Mergias-Robles et al., 2022; Parker & Morgan, 2020), thus preventing their translation into criminal behaviours.

Therefore, the present study focuses on multiple psychological predictors of aggression, violent behaviour (in violent offenders), and attitudes towards lying as a proxy for deception, across two distinct populations: violent offenders and the general population. We will first review essential individual traits and group differences between violent offenders and the general population, followed by a literature review of the relation between these factors and aggression, violent behaviour, and attitudes towards lying. So far, the limited research on aggression and deception in legal and forensic settings has looked less at psychological and individual determinants compared to clinical and situational factors (Steinert, 2002). Additionally, most existing studies typically examined only one or two predictors at a time, rather than investigating multiple individual predictors simultaneously (Komasi et al., 2016). Aggression, violent behaviour, and deception are multifaceted constructs, and examining them from a single perspective (e.g., as personality traits) provides an incomplete picture; thus, it is essential to consider models that incorporate multiple factors to gain a more comprehensive understanding (Komasi et al., 2016). Hence, the present study will further investigate the combined predictive power of individual differences in personality traits (FFM), sensitivity to reward and punishment (BAS Drive, BAS Fun Seeking, BAS Reward Responsibility, BIS), Callous-Unemotional (CU) traits (callousness, unemotional, uncaring), social desirability (self-deceptive enhancement, SDE and impression management, IM) and socio-demographic variables (age, gender, level of education) as possible predictors for both aggression, violent acts committed in prison (in the case of violent offenders sample only) and attitudes towards lying (see Figure 1 for a visual representation of all predictors). The present study aims to confirm previous literature findings while also exploring new associations by examining models with multiple predictors. In terms of outcomes, we will look at *self-reported* aggression as an overall measure and its components, such as anger, verbal and physical aggression and hostility (Buss & Perry, 1992). Additionally, for the violent offenders' sample, we included a more *objective* measure of violent behaviour (prison records of the violent acts committed in prison).

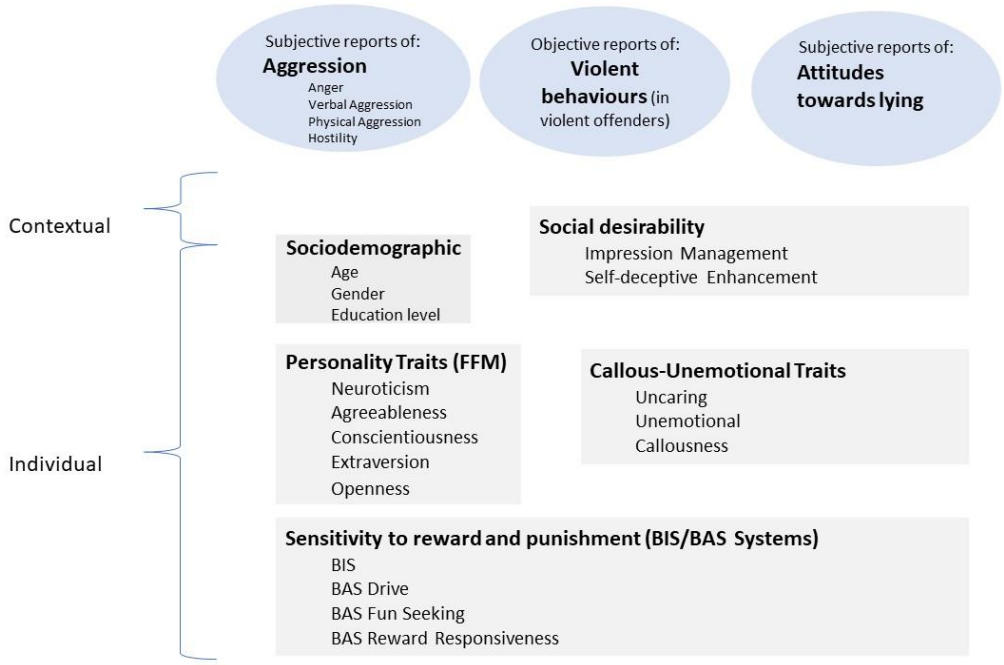
First, we looked at group differences, expecting higher levels of undesirable behaviours in terms of *aggression*, components of aggression (Confirmatory hypothesis 1, C1) and *attitudes towards lying* (C2) in violent offenders compared with the general population. Moreover, some differences were also expected in *personality* traits, such as higher

Neuroticism, lower Agreeableness and lower Conscientiousness in violent offenders compared with the general population (C3a). At the same time, Openness and Extraversion might be lower in violent offenders, or there might not be a significant difference, as studies show less consistent results for these traits (C3b). Additionally, we expected to find significant differences in *sensitivity to reward and punishment* with BAS (and BAS subfactors, respectively) at increased levels and BIS at decreased levels for violent offenders compared with the general population (C4). Although no studies, to our knowledge, looked at the group differences in CU traits in adults, we expected CU traits to be increased in the violent offenders' sample compared with the general population (Exploratory hypothesis 1, E1). We don't have conceptual grounds to expect to find significant differences between these populations in self-deceptive enhancement (E2a). However, impression management might be higher in violent offenders than in the general population, as several studies pointed out its connection with criminal risk variance (E2b). Next, when attempting to predict undesirable behaviours in the two groups, we expected high Neuroticism, lower Agreeableness and lower Conscientiousness to be associated with both higher self-reported aggression and its components (C6), with objective violent behaviour (E3) and with more lenient attitudes towards lying (E4).

Lastly, using our self-reported measures to create a composite model to predict our outcomes (aggression and its subcomponents, violent behaviour, and attitudes towards lying), we explored the added predictive value of subfactors of sensitivity to reward and punishment (BAS Reward Responsiveness, BAS Fun Seeking, BAS Drive and BIS), CU trait subfactors (uncaring, callousness and unemotional), social desirability (SDE and IM), socio-demographic variables (age, gender and educational level). This model is rather exploratory (E5) and can significantly extend the literature on individual predictors of the most aversive behaviours characterising criminal conduct (aggression, violent behaviour and deception). Furthermore, by investigating these predictors across diverse populations (violent offenders – E5a and the general population- E5b), our study aims to provide valuable insights into the underlying mechanisms of aggression, violent behaviour and deceptive behaviours (the "seeds of trouble" as suggested by our title).

Figure 1

Individual (and Contextual) Predictors of Aggression and its Components, Violent Behaviour and Attitudes Towards Lying



3.4.1 Methods

Data is available at https://osf.io/zhgb4/.

General population. 238 participants from the general population took part in this study (M_{age}= 36,69, SD_{age} =10.65, ranging from 20 to 68 years old; 47.5% men (n = 115), and 52.1% women (n = 125). 5.9% of the participants reported a lower educational level (i.e., completed between 4 and 8 grades; n =14), and the rest of 92.9% (n = 221) reported a higher educational level (i.e., high school; theoretical or vocational and higher degrees).

Violent offenders. The prison sample included 202 inmates incarcerated for violent offences (see Table 1) (M_{age}= 38,30; SD_{age} =10.80, ranging from 21 to 69 years old, 84.1% men (n = 171), and the rest of 15.3% women (n = 31). The participants were recruited from two maximum-security, mixed prisons, prisons with 70.3% of the inmates being in a closed regime, 19.3 % in a maximum-security regime and the remaining 10.4% in a semi-open or open regime. The maximum sentence duration is a life sentence with the possibility of parole, while the minimum is 10 months. A total of 44.6% of the participants reported a lower educational background (n = 90), while the rest (55.35 %, n = 112) completed higher education. During the time spent in prison, 22.9 % of the offenders committed violent behaviour as violent acts

towards others (e.g., one-to-one fights, see Measures section for a detailed description; *n* =53), with the number of violent acts committed varying from 1 (12.4%, *n* =25) to 16 (0.5%, *n* =1).

Table 1

Frequency and Percentage of the Type of Offence in the Violent Offenders' Sample

Type of offence	*n*	%
Murder or serious injury (murder, manslaughter, homicide, assault)	156	77.2
Sex-related crimes (sexual assault, rape, child molestation, sexual activity with a child)	18	8.9
Other crimes (firearm possession, blackmail, kidnapping, human trafficking, organised crime, pimping).	6	3.0
Theft or robbery (theft, burglary, armed robbery and robbery; for each of these offenders, the description of the offence included observable violent acts)	22	10.9

Note. n = 202.

Self-reported Aggression and Components of Aggression. The Aggression Questionnaire (AQ: Buss & Perry, 1992) measures a total aggression score (i.e., a higher score means higher aggression) and four components (physical and verbal aggression, anger and hostility). Total aggression Cronbach's Alpha was α = .88 for violent offenders, respectively .86 for the general population. Moreover, when looking at components of aggression subscales separately, Cronbach's Alpha was α = .66 for violent offenders and α = .68 for the general population on Anger; α = .47 for violent offenders and α = .49 for the general population on Verbal Aggression; α = .79 for violent offenders and α = .78 for the general population on Physical Aggression; and α = .71 for violent offenders and α = .69 for the general population on Hostility.

Violent behaviour consists of violent acts committed in prison**.** This measure applies exclusively to inmates and serves as an objective assessment of violent behaviour committed by prisoners during incarceration. These behaviours are investigated by the prison board and documented by the prison officers on duty.

Attitudes towards lying (lie acceptability) were measured using the Revised Lie Acceptability Scale (RLAS; Oliveira & Levine, 2008). The scale measures the individual-level variability in attitudes toward deception across 11 five-point Likert scale items ("It is ok to lie to achieve one's goals."), scored from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). Higher scores reflect higher levels of lie acceptability or more favourable attitudes towards lying. Cronbach's Alpha was α = .62 for the violent offenders and α = .69 for the general population.

Personality traits were measured using the NEO Personality Inventory–Revised-3 (NEO-PI-R-3; McCrae & Costa, 2010; Iliescu & Sârbu, 2019), which measures a comprehensive

model of general personality traits, the Five-Factor Model or "Big Five": Agreeableness, Conscientiousness, Neuroticism, Extraversion, and Openness to Experience. In this study, Cronbach's Alpha was α = .67 for the violent offenders and α = .68 for the general population (see Supplemental materials for Cronbach's Alpha on each subscale).

Sensitivity to reward and punishment was measured as self-reports using the BIS/BAS Scales (Carver & White, 1994). Cronbach's Alpha was α = .62 for the violent offenders and α = .60 for the general population for BAS Drive, α = .54 for the violent offenders and α = .59 for the general population for BAS Fun Seeking, α = .73 for the violent offenders and α = .61 for the general population for BAS Reward Responsiveness and α = .37 for the violent offenders and α = .65 for the general population for BIS.

Callous-unemotional traits were measured using the Inventory of Callous–Unemotional Traits (ICU; Frick, 2004; Roșan et al., 2015). Cronbach's Alpha was α = .65 for the violent offenders and α = .80 for the general population for Uncaring, α = .64 for the violent offenders and α = .69 for the general population for Callousness, and α = .48 for the violent offenders and α = .62 for the general population for Unemotional Scale.

Social desirability, more specifically, self-deceptive enhancement and impression management, was measured using the Paulhus Deception Scales (PDS; Paulhus, 1998; an updated version of the Balanced Inventory of Desirable Responding, BIDR; Paulhus, 1991). The scales can be scored as continuous or dichotomous, with continuous scoring (used in this study) methods, as supported by Stöber et al. (2002) and Schnapp et al. (2017), showing higher Cronbach's alphas, convergent correlations with social desirability measures and better reliability. The PDS was also validated in the general population and with prison inmates (Lanyon & Carle, 2007). For this study, Cronbach's alpha was for SDE α = .40 in violent offenders and α = .56 in the general population, and for IM it was α = .73 for the violent offenders and α = .79 in the general population.

This study employed a quantitative, correlational design, utilising hierarchical multiple regression as the primary analysis procedure. The aim was to investigate the influence of multiple independent variables on dependent variables. We constructed a series of nested regression models, where each subsequent model incorporates all variables from the previous model plus additional predictors, thus investigating how new variables impact the overall model and explain additional variance in the outcome. Thus, we examined the incremental changes in explained variance as new predictors are added to the model while controlling for previously entered variables (e.g., does sensitivity to reward and punishment add variance to aggression when personality is controlled). The order of variable entry is based on theoretical

considerations, as detailed in the Introduction section (i.e., personality added in step one as it is the most robust and consistent finding in the literature on predictors of aggression; Jones et al., 2011).

3.4.2 Results and Conclusions

Due to item nonresponse and incomplete data (i.e., the total percentage of missing data values per participant > 25%), we included in the final analysis data from $N = 440$ participants ($n = 202$ violent offenders and $n = 238$ general population). The remaining missing data was less than 5% in total for all participants; however, multiple imputation (pooled using the "Bar Procedure"; Baranzini, 2015) was used for missing data replacement due to the complex statistical analysis done in this study (Eekhout et al., 2014).

Descriptive Statistics and Group Differences

The means and standard deviations of all variables for the general population and the violent offenders' samples are reported in Table 2. All results are considered significant at an alpha level of $\alpha = .0025$ as a Bonferroni correction was used to reduce the risk of Type I errors due to multiple comparisons (Dunn, 1961).

Table 2

Descriptive Statistics and Group Differences

	Violent offenders ($N = 202$)		General population ($N = 238$)		*t*	*p*	Cohen's *d*	95% CI	
	M	*SD*	*M*	*SD*					
Aggression	66.95	17.93	55.80	14.48	**7.09*****	< .001	0.69	8.06	14.24
Anger	15.45	5.18	13.39	4.53	**4.04*****	< .001	0.42	1.14	2.98
Verbal aggression	12.28	3.38	11.05	3.03	**4.02*****	< .001	0.38	0.63	1.83
Physical aggression	18.76	7.34	14.16	5.82	**7.18*****	< .001	0.74	3.33	5.85
Hostility	20.46	5.80	17.19	4.96	**6.30*****	< .001	0.60	2.25	4.29
Lie acceptability	28.25	7.20	25.44	6.86	**4.18*****	< .001	0.40	1.49	4.13
Neuroticism	34.80	6.89	32.22	8.11	**3.61*****	< .001	0.34	1.18	3.99
Extraversion	40.12	6.35	43.62	6.49	**-5.69*****	< .001	-0.54	-4.71	-2.29
Openness	41.34	5.73	42.29	5.34	-1.80	.07	-0.17	1.99	0.09
Agreeableness	42.70	7.00	45.14	6.02	**-3.95*****	< .001	-0.37	-3.67	-1.23
Conscientiousness	47.62	7.32	48.56	6.63	-1.41	.16	-0.13	-2.24	0.37
BAS Drive	11.59	2.67	11.66	2.06	- 0.31	.76	-0.03	-0.52	0.37
BAS Fun Seeking	11.06	2.47	11.19	2.33	- 0.54	.58	-0.05	-0.57	0.32
BAS Reward Responsiveness	16.42	2.88	17.37	1.99	**-3.99*****	< .001	-0.39	-1.43	-0.49
BIS	18.75	2.98	19.57	3.27	-2.72	.007	-0.26	-1.41	-0.23
Uncaring	13.68	4.62	18.20	4.57	**-10.29*****	< .001	-0.98	-5.39	-3.66
Callousness	26.91	7.45	20.75	5.35	**9.81*****	< .001	0.96	4.93	7.40

	Violent offenders (N = 202)		General population (N = 238)		t	p	Cohen's d	95% CI	
Unemotional	13.59	3.93	14.21	3.50	-1.75	.08	0.17	-1.31	0.75
Self-Deceptive Enhancement[12]	66.05	7.60	67.90	8.03	-2.46	.01	-0.23	-3.32	-0.37
Impression Management[13]	65.34	12.05	70.27	12.01	**-4.28*****	< .001	0.40	-7.19	-2.67
Age	38.30	10.81	36.69	10.65	1.57	.11	0.15	-0.41	3.62

Note. CI = Confidence Interval; ** p < .0025, ****p* < .001.

Results show that the violent offender sample reported statistically significantly higher aggression (both as a total score, as well as per each component: anger, verbal and physical aggression and hostility) and more favourable attitudes towards lying, therefore confirming *C1 and C2*.

Also, violent offenders reported higher Neuroticism, lower Agreeableness (partially confirming *C3a*; no statistically significant difference in Conscientiousness) and lower Extraversion (partially confirming *C3b*; no statistically significant difference in Openness). Regarding the sensitivity to reward and punishment, only BAS Reward Responsiveness was lower in violent offenders (results did not support *C4)*; for BIS, it was reduced as expected; however, the differences between groups were not statistically significant. Uncaring was significantly higher in the general population compared with violent offenders, while callousness was higher for violent offenders, as expected (*E1*). Impression management was higher in the general population than in violent offenders (*E2b* was not supported), and a not statistically significant but reduced self-deceptive enhancement was reported by violent offenders compared to the general population (*E2a)*. There was no statistically significant difference in age. Effect sizes for all statistical differences varied from Cohen's d = .34 to d = .98.

Predictors of Undesirable Behaviours in Violent Offenders and the General Population

We were interested in building models that used all relevant variables and best predicted our outcomes (Lewis, 2007) while avoiding overfitting the data. In this regard, we checked that the ratio of observations per predictor variable falls within the 10 to 15 observations per variable's rule of thumb (Althubaiti, 2023), the normality of homoscedasticity and the normality of residuals.

Aggression

[12] SDE as a dichotomous score: Violent Offenders M = 5.71, SD = 4.11; General population M = 5.37 SD = 3.78.
[13] IM as a dichotomous score: Violent Offenders M = 9.75, SD = 3.83; General population M = 11.36, SD = 4.01.

A multiple hierarchical regression analysis was conducted to examine the predictors of aggression in violent offenders and the general population (*testing C6 and E4*; see Table 3 for a summary of results and steps, and Supplementary materials for a detailed report). The full model was statistically significant for both groups, with an effect of $f^2 = 0.85$ for violent offenders and $f^2 = 0.98$. It accounted for 46% and 49% of the variance in aggression for violent offenders and the general population, respectively. A combination of high Neuroticism, low Agreeableness, high BAS Fun Seeking and BAS Drive, high callousness, low self-deceptive enhancement, and low impression management accounted for a substantial portion of the variance in aggression for violent offenders and the general population, with slight differences in predictors for each population.

Components of Aggression

Agreeableness was found to be a statistically significant positive predictor of anger, verbal aggression, physical aggression, and hostility in both violent offenders and the general population. Neuroticism was a statistically significant negative predictor in both samples for anger, verbal aggression and hostility. Conscientiousness was a statistically significant negative predictor for anger. BAS Fun Seeking in violent offenders and BAS Drive in the general population were found to be statistically significant positive predictors of physical aggression. Callousness was a statistically significant negative predictor of physical aggression in both samples. Impression management was found to be a statistically significant negative predictor of anger and physical aggression for both samples and hostility in the general population. Self-deceptive enhancement was found to be a statistically significant negative predictor of anger and physical aggression in violent offenders. The effect sizes for the models ranged from $f^2 = 0.18$ to $f^2 = 0.86$. The models accounted for 15% to 47% of the variance in the outcome variables.

Table 3

Hierarchical Regression for Aggression in Violent Offenders on the left (n = 202) and the General Population on the Right (n = 238)

Variable	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2
Step 1						0.40	0.41***						0.45	0.45***
Constant	85.96***	62.64	109.29	11.83				94.52***	78.10	110.94	8.33			
Neuroticism	0.96	0.66	1.26	0.15	.37***			0.63	0.45	0.82	0.09	.35***		
Openness	-0.20	-0.55	0.16	0.18	-.06									
Agreeableness	-1.04	-1.33	-0.74	0.15	-.40***			-1.12	-1.36	-0.88	0.12	-.46***		
Conscientiousness								-0.18	-0.41	0.05	0.12	-.08		
Step 2						0.42	0.02*						0.47	0.03**
Constant	80.51***	57.11	103.92	11.87				78.60***	60.19	97.01	9.34			
Neuroticism	0.87	0.57	1.18	0.02	.34***			0.60	0.42	0.78	0.09	.33***		
Openness	-0.27	-0.63	0.08	0.18	-.09									
Agreeableness	-1.03	-1.32	-0.74	0.15	-.40***			-1.01	-1.26	-0.76	0.13	-.42***		
Conscientiousness								-0.24	-0.48	-0.01	0.12	-.11*		
BAS D								0.86	0.08	1.65	0.40	.12*		
BAS FS	1.02	0.22	1.82	0.40	.14*			0.48	-0.18	1.14	0.33	.08		
Step 3						0.43	0.01*						.48	0.01
Constant	67.65***	41.29	94.02	13.37				77.50***	51.41	103.59	13.24			
Neuroticism	0.81	0.50	1.11	0.16	.31***			0.57	0.39	0.75	0.09	.32***		
Openness	-0.19	-0.55	0.17	0.18	-.06									
Agreeableness	-0.94	-1.24	-0.65	0.15	-.37***			-1.02	-1.32	-0.72	0.15	-.42***		
Conscientiousness								-0.22	-0.46	0.01	0.12	-.10		
BAS D								0.81	0.03	1.59	0.40	.12*		
BAS FS	1.02	0.23	1.82	0.40	.14*			0.48	-0.18	1.14	0.34	.08		
Uncaring								-0.17	-0.52	0.18	0.18	-.05		
Callousness	0.29	0.01	0.59	0.15	.12*			0.25	-0.03	0.52	0.14	.09		
Step 4						0.46	0.04**						.49	0.02**
Constant	95.26***	64.86	125.66	15.41				88.32***	59.54	117.10	14.60			
Neuroticism	0.70	0.40	1.01	0.15	.27***			0.51	0.31	0.71	0.10	.29***		
Openness	-0.01	-0.37	0.36	0.19	-.00									
Agreeableness	-0.82	-1.12	-0.51	0.15	-.32***			-0.92	-1.22	-0.62	0.15	-.38***		
Conscientiousness								-0.09	-0.34	0.15	0.13	-.04		
BAS D								0.80	0.04	1.57	0.39	.11*		
BAS FS	0.80	0.02	1.59	0.39	.11*			0.37	-0.29	1.02	0.33	.06		
Uncaring								-0.15	-0.50	0.20	0.18	-.05		
Callousness	0.28	0.001	0.57	0.14	.11*			0.24	-0.03	0.51	0.14	.09		
SDE	-0.30	-0.57	-0.03	0.14	-.13*			-0.07	-0.29	0.15	0.11	-.04		
IM	-0.22	-0.40	-0.04	0.09	-.15*			-0.20	-0.33	-0.06	0.07	-.16**		
	F (7, 194) = 25.50; p < .001; f^2 = 0.85							F (9, 228) = 26.80; p < .001; f^2 = 0.98						

Note. CI = confidence interval; *p < .05; **p < .01; ***p < .001

Table 4

Hierarchical Regression for Anger in Violent Offenders (n = 202) and the General Population (n = 238)

Variable		*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2
Step 1							0.33	0.34***						0.34	0.34***
	Constant	14.49***	6.95	22.04	3.82				20.856***	15.22	26.50	2.86			
	Neuroticism	0.33	0.23	0.42	0.05	.44***			0.21	0.15	0.27	0.03	.38***		
	Openness	-0.12	-0.23	0.00	0.06	-.13									
	Agreeableness	-0.19	-0.28	-0.10	0.05	-.26***			-0.23	-0.31	-0.15	0.04	-.30***		
	Conscientiousness	0.05	-0.04	0.15	0.05	.08			-0.08	-0.16	-0.01	0.04	-.12*		
Step 2														0.35	0.02*
	Constant								16.67***	10.28	23.05	3.24			
	Neuroticism								0.20	0.14	0.26	0.03	.36***		
	Openness														
	Agreeableness								-0.20	-0.29	-0.12	0.04	-.27***		
	Conscientiousness								-0.09	-0.17	-0.01	0.04	-.14*		
	BAS Drive								0.17	-0.10	0.44	0.14	.08		
	BAS FS								0.18	-0.05	0.41	0.12	.09		
Step 2/3							.34	.01						0.35	0.00
	Constant	10.59**	2.02	19.16	4.34				15.81***	8.65	22.98	3.64			
	Neuroticism	0.31	0.22	0.41	0.05	.41***			0.20	0.14	0.26	0.03	.36***		
	Openness	-0.10	-0.22	0.02	0.06	-.11									
	Agreeableness	-0.17	-0.26	-0.08	0.05	-.23***			-0.20	-0.29	-0.11	0.05	-.26***		
	Conscientiousness	0.06	-0.03	0.16	0.05	.09			-0.09	-0.17	-0.01	0.04	-.13*		
	BAS Drive								0.17	-0.11	0.44	0.14	.08		
	BAS FS								0.18	-0.05	0.41	0.12	.09		
	Callousness	0.09	-0.01	0.18	0.05	.12			0.03	-0.07	0.12	0.05	.03		
Step 3/4							.38	.05**						0.36	0.01
	Constant	18.60***	9.17	28.03	4.78				20.81***	12.40	29.22	4.27			
	Neuroticism	0.27	0.18	0.37	0.05	.36***			0.17	0.09	0.24	0.04	.30***		
	Openness	-0.06	-0.17	0.06	0.06	-.06									
	Agreeableness	-0.14	-0.23	-0.04	0.05	-.19**			-0.18	-0.28	-0.09	0.05	-.24***		
	Conscientiousness	0.10	-0.00	0.19	0.05	.14			-0.05	-0.14	0.03	0.04	-.08		
	BAS Drive								0.17	-0.10	0.44	0.14	.08		
	BAS FS								0.16	-0.07	0.39	0.12	.08		
	Callousness	0.09	-0.00	0.17	0.04	.12			0.03	-0.06	0.13	0.05	.04		
	SDE	-0.11	-0.19	-0.02	0.04	-.16*			-0.06	-0.14	0.01	0.04	-.11		
	IM	-0.07	-0.12	-0.01	0.03	-.15*			-0.03	-0.08	0.02	0.02	-.07		
		F (7, 194) = 18.48; p < .001; f^2 = 0.60							F (8, 229) = 17.58; p < .001; f^2 = 0.56						

Note. CI = confidence interval; *p < .05; **p < .01; ***p < .001

Table 5

Hierarchical Regression for Verbal Aggression in Violent Offenders (n = 202) and the General Population (n = 238)

Variable	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2
Step 1						0.19	0.20***						0.15	0.16***
Constant	15.38***	11.40	19.35	2.02				17.60***	14.27	20.93	1.69			
Neuroticism	0.11	0.04	0.17	0.03	.22**			0.05	0.00	0.09	0.02	.13*		
Agreeableness	-0.16	-0.22	-0.10	0.03	-.33***			-0.18	-0.24	-0.12	0.03	-.36***		
Step 2						0.19	0.00						0.16	0.01
Constant	14.47**	10.03	18.92	2.25				15.86***	11.90	19.82	2.01			
Neuroticism	0.10	-0.03	0.17	0.03	.20**			0.04	-0.00	0.09	0.02	.12		
Agreeableness	-0.15	-0.22	-0.08	0.03	-.31***			-0.16	-0.23	-0.10	0.03	-.32***		
Callousness	0.03	-0.03	0.09	0.03	.06			0.06	-0.01	0.13	0.04	.10		
Step 3													0.16	0.00
Constant								16.46***	12.11	20.81	2.21			
Neuroticism								0.04	-0.01	0.09	0.02	.11		
Agreeableness								-0.16	-0.22	-0.09	0.03	-.31***		
Callousness								0.06	-0.02	0.13	0.04	.10		
IM								-0.01	-0.04	0.02	0.02	-.04		
						$F(3, 198) = 16.65$, $p < .001$, $f^2 = 0.23$							$F(4, 233) = 12.12$, $p < .001$, $f^2 = 0.18$	

Note. CI = confidence interval; *$p < .05$; **$p < .01$; ***$p < .001$.

Table 6

Hierarchical Regression for Physical Aggression in Violent Offenders (n = 202) and the General Population (n = 238)

Variable	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2
Step 1						0.34	0.35***						0.30	0.31***
Constant	41.08***	30.48	51.68	5.37				32.65***	25.23	40.07	3.76			
Neuroticism	0.17	0.04	0.30	0.07	.16*			0.14	0.05	0.22	0.04	.19**		
Openness	-0.21	-0.37	-0.04	0.09	-.16*									
Agreeableness	-0.51	-0.64	-0.38	0.07	-.48***			-0.44	-0.55	-0.33	0.06	-.45***		
Conscientiousness	0.04	-0.10	0.18	0.07	.04			-0.06	-0.16	0.04	0.05	-.07		
Step 2						0.38	0.04***						0.322	0.023**
Constant	38.42***	28.04	48.80	5.27				27.34***	19.15	35.53	4.16			
Neuroticism	0.11	-0.02	0.24	0.07	.10			0.12	0.04	0.20	0.04	.17**		
Openness	-0.24	-0.40	-0.08	0.08	-.19**									
Agreeableness	-0.50	-0.62	-0.37	0.06	-.47***			-0.39	-0.51	-0.28	0.06	-.41***		
Conscientiousness	0.01	-0.12	0.15	0.06	.01			-0.10	-0.20	0.01	0.05	-.11		
BAS Drive								0.45	0.14	0.77	0.16	.16**		
BAS FS	0.63	0.29	0.97	0.17	.21***									
Step 3						0.39	0.02						0.34	0.02*
Constant	30.50***	17.98	43.02	6.35				26.59***	15.10	38.08	5.83			
Neuroticism	0.09	-0.04	0.22	0.07	.08			0.11	0.03	0.19	0.04	.15**		
Openness	-0.21	-0.37	-0.04	0.08	.16*									
Agreeableness	-0.45	-0.58	-0.32	0.07	-.43***			-0.40	-0.53	-0.26	0.07	-.41***		
Conscientiousness	0.04	-0.10	0.18	0.07	.04			-0.08	-0.19	0.02	0.05	-.10		
BAS Drive								0.43	0.11	0.74	0.16	.15**		
BAS FS	0.62	0.29	0.96	0.17	.21***									
Uncaring	0.09	-0.10	0.27	0.09	.06			-0.09	-0.25	0.07	0.08	-.07		
Callousness	0.12	-0.00	0.25	0.06	.12			0.14	0.01	0.26	0.06	.13*		
Step 4						0.47	0.08***						0.36	0.03*
Constant	48.38***	34.72	62.04	6.92				27.23*	14.46	40.01	6.48			
Neuroticism	0.03	-0.09	0.16	0.06	.03			0.11	0.02	0.20	0.05	.15*		
Openness	-0.13	-0.29	0.03	0.08	-.10									
Agreeableness	-0.40	-0.52	-0.27	0.06	-.38***			-0.35	-0.48	-0.21	0.07	-.36***		
Conscientiousness	0.09	-0.04	0.22	0.07	-.09			-0.04	-0.15	0.07	0.06	-.05		
BAS Drive								0.39	0.08	0.70	0.16	.14*		
BAS FS	0.47	0.15	0.79	0.16	.16**									
Uncaring	-0.02	-0.19	0.16	0.09	-.01			-0.08	-0.23	0.08	0.08	-.06		
Callousness	0.13	0.01	0.24	0.06	.13*			0.13	0.00	0.25	0.06	.12*		
SDE	-0.18	-0.29	-0.07	0.06	-.19**			0.03	-0.07	0.13	0.05	.04		
IM	-0.14	-0.22	-0.07	0.04	-.23***			-0.10	-0.16	-0.03	0.03	-.20**		
	$F(9, 192) = 20.43$, $p < .001$, $f^2 = 0.86$							$F(8,229) = 17.43$, $p < .001$, $f^2 = 0.55$						

Note. CI = confidence interval; *p < .05; **p < .01; ***p < .001

Table 7

Hierarchical Regression for Hostility in Violent Offenders (n = 202) and the General Population (n = 238)

Variable		*B*	95% CI for B		SE B	Beta	Adj R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adi R^2	ΔR^2
Step 1							0.28	0.29***						0.322	0.33***
	Constant	15.85***	9.42	22.28	3.26				24.80***	17.80	31.80	3.56			
	Neuroticism	0.36	0.26	0.46	0.05	.43***			0.23	0.16	0.30	0.04	.38***		
	Extraversion								-0.04	-0.13	0.04	0.04	-.06		
	Agreeableness	-0.19	-0.29	-0.08	0.05	-.22***			-0.26	-0.36	-0.17	0.05	-.32***		
	Conscientiousness								-0.02	-0.11	0.06	0.05	-.03		
Step 2														0.35	0.04**
	Constant								19.42***	11.82	27.03	3.86			
	Neuroticism								0.19	0.11	0.26	0.04	.30***		
	Extraversion								-0.09	-0.18	-0.00	0.05	-.12*		
	Agreeableness								-0.25	-0.34	-0.14	0.05	-.30***		
	Conscientiousness								-0.03	-0.12	0.06	0.05	-.04		
	BAS Drive								0.19	-0.11	0.49	0.15	.08		
	BAS FS								0.29	0.03	0.56	0.13	.14*		
	BIS								0.14	-0.04	0.31	0.09	.09		
Step 2/3							0.28	0.00						0.35	0.00
	Constant	14.44***	7.25	21.62	3.64				14.70*	3.95	25.46	5.46			
	Neuroticism	0.35	0.24	0.46	0.05	.42***			0.18	0.11	0.26	0.04	.30***		
	Extraversion								-0.09	-0.18	0.00	0.05	-.12*		
	Agreeableness	-0.17	-0.28	-0.06	0.05	-.21**			-0.21	-0.32	-0.09	0.06	-.25***		
	Conscientiousness								-0.02	-0.11	0.07	0.05	-.02		
	BAS Drive								0.19	-0.11	0.49	0.15	.08		
	BAS FS								0.31	0.04	0.58	0.13	.15*		
	BIS								0.16	-0.03	0.34	0.09	.10		
	Uncaring								0.08	-0.06	0.22	0.07	.07		
	Callousness	0.04	-0.06	0.15	0.05	.06			0.03	-0.07	0.14	0.05	.04		
Step 3/4							0.28	0.01						0.37	0.02*
	Constant	19.49**	9.20	29.78	5.22				18.85**	6.73	30.97	6.15			
	Neuroticism	0.33	0.22	0.44	0.06	.39***			0.17	0.09	0.25	0.04	.27***		
	Extraversion								-0.09	-0.18	0.00	0.05	-.12		
	Agreeableness	-0.16	-0.27	-0.05	0.06	-.19**			-0.17	-0.29	-0.06	0.06	-.21**		
	Conscientiousness								0.02	-0.07	0.12	0.05	.03		
	BAS Drive								0.17	-0.12	0.47	0.15	.07		
	BAS FS								0.27	0.01	0.54	0.13	.13*		
	BIS								0.13	-0.06	0.31	0.09	.08		
	Uncaring								0.08	-0.06	0.22	0.07	.07		
	Callousness	0.04	-0.06	0.14	0.05	.05			0.03	-0.07	0.14	0.05	.03		
	SDE	-0.06	-0.15	0.04	0.05	-.07			-0.03	-0.11	0.06	0.04	-.04		

Variable	*B*	95% CI for B		SE B	Beta	Adj R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adi R^2	ΔR^2
IM	-0.02	-0.08	0.05	0.03	-.03			-0.06	-0.11	-0.01	0.03	-.15*		
						$F(5, 196) = 16.78$, $p < .001$, $f^2 = 0.39$							$F(11,226) = 13.42$, $p < .001$, $f^2 = 0.52$	

Note. CI = confidence interval; *$p < .05$; **$p < .01$; ***$p < .001$.

Violent Behaviour

A logistic regression analysis was conducted and the model was statistically significant (see Table 8 for a summary of results) and explained 9.7% (Nagelkerke R^2 = 0.97) of the variance in violent behaviour in prison, with age and gender (men) being significant predictors of the likelihood of committing violent acts. Age was a statistically significant predictor in the model, with every one-year increase in age, the odds of engaging in violent acts decreased by 4% ((1-0.96)*100). Gender (women vs. men) was also a statistically significant predictor, with women having 4.87% lower odds of engaging in violent acts compared to men.

Table 8

Logistic Regression for Violent Behaviour (yes/no) in Violent Offenders (n = 202)

Predictor Variables	B	SE	Wald	df	*p*-value	Odds Ratio	95% CI for Odds Ratio
Constant	-1.06	0.96	1.22	1	.269	0.34	
Age	-0.04	0.02	5.91*	1	.015	0.96	[0.93, 0.99]
Gender (Men vs. Women)	1.58	0.76	4.36*	1	.037	4.87	[1.10, 21.48]
Hosmer-Lemeshow Test			$\chi 2(2)=13.29$	8	.231		

Note. CI = confidence interval; *$p < .05$; **$p < .01$; ***$p < .001$.

Attitudes Towards Lying

A hierarchical regression analysis was conducted to examine the predictors of attitudes towards lying in violent offenders and the general population (testing *C4, E5a, E5b*; see Table 9 for a summary of results). The final step of the model was statistically significant with an effect size of f^2 = 0.50, respectively f^2 = 0.56, accounting for 34% of the variance in violent offenders and 35% in the general population. Neuroticism and impression management were statistically significant positive respectively negative predictors in both groups, while callousness was a statistically significant positive predictor only in violent offenders.

Table 9

Hierarchical Regression for Attitudes Towards Lying in Violent Offenders (n = 202) and the General Population (n = 238)

Variable		*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2	*B*	95% CI for B		SE B	Beta	Adjusted R^2	ΔR^2
Step 1							0.27	0.29***						0.22	0.23***
	Constant	42.90***	31.35	54.44	5.85				35.54***	26.31	44.76	4.68			
	Neuroticism	0.26	0.13	0.40	0.07	.25***			0.24	0.14	0.35	0.05	0.29***		
	Extraversion	0.10	-0.15	0.17	0.08	.01									
	Openness	-0.15	-0.33	0.02	0.09	-.12									
	Agreeableness	-0.25	-0.39	-0.12	0.07	-.24***			-0.36	-0.50	-0.23	0.07	-.32***		
	Conscientiousness	-0.15	-0.30	0.01	0.08	-.15			-0.03	-0.16	0.10	0.07	-.03		
Step 2							0.30	0.03**						0.25	0.03*
	Constant	33.59***	20.79	46.38	6.49				22.70***	7.97	35.57	7.26			
	Neuroticism	0.22	0.09	0.36	0.07	.21**			0.23	0.13	0.34	0.05	.28***		
	Extraversion	0.01	-0.14	0.17	0.08	.01									
	Openness	-0.11	-0.28	0.06	0.09	-.09									
	Agreeableness	-0.20	-0.34	-0.06	0.07	-.20**			-0.27	-0.43	-0.10	0.08	-.23**		
	Conscientiousness	-0.13	-0.28	0.02	0.08	-.13			-0.01	-0.13	-0.13	0.07	-.00		
	Uncaring								0.06	-0.15	0.26	0.10	.04		
	Callousness	0.20	0.07	0.33	0.07	.21**			0.20	0.04	0.36	0.08	.16*		
	Unemotional								0.17	-0.05	0.40	0.17	.12		
Step 3							0.34	0.04**						0.35	0.11***
	Constant	37.92**	25.19	50.65	6.46				27.34**	12.20	42.49	7.68			
	Neuroticism	0.20	0.07	0.34	0.07	.20**			0.20	0.10	0.31	0.06	.24***		
	Extraversion	0.04	-0.12	0.19	0.08	.03									
	Openness	-0.08	-0.25	0.09	0.09	-.06									
	Agreeableness	-0.13	-0.27	0.00	0.07	-.13			-0.14	-0.30	0.02	0.08	-.12		
	Conscientiousness	-0.12	-0.27	0.03	0.07	-.12			0.11	-0.02	0.24	0.07	.11		
	Uncaring								0.09	-0.09	0.28	0.10	.06		
	Callousness	0.19	0.06	0.31	0.06	.19**			0.18	0.04	0.33	0.08	.15		
	Unemotional								0.10	-0.12	0.31	0.10	.05		
	SDE								0.02	-0.10	0.14	0.06	.02		
	IM	-0.13	-0.21	-0.05	0.04	-.22**			-0.23	-0.30	-0.15	0.04	-.40***		
						$F(7, 194) = 15.67$, $p < .001$, $f^2 = 0.50$							$F(8, 229) = 17.05$, $p < .001$, $f^2 = 0.56$		

Note. CI = confidence interval; *$p < .05$; **$p < .01$; ***$p < .001$.

The two aims of this study were to investigate potential differences in aggression and attitudes towards lying, along with identifying the individual predictors of these phenomena in the general population and violent offenders. Understanding the similarities and differences between the two populations and identifying the individual predictors for aggression, violent behaviour, and attitudes towards lying is essential for developing targeted prevention programs and interventions for both offenders and the general public. Moreover, it is useful for assessment measures, especially when legal decisions are made based on these, such as risk assessment (Steinert, 2002; Komasi et al., 2016; Mergias-Robles et al., 2022; Parker & Morgan, 2020). The results of our study confirmed that violent offenders exhibit higher levels of trait aggression and its components, meaning anger, verbal and physical aggression, and hostility, and report more positive attitudes towards lying compared to the general population. These findings are consistent with prior research indicating that aggression is more prevalent in legal and forensic contexts (Hornsveld & Kraaimaat, 2023; Ohlsson & Ireland, 2011) and that knowledge about deception and attitudes towards deception differ between the prisoners and non-offenders (e.g., Granhag et al., 2010; Hurezan et al., 2024).

Extending the limited literature in the field on this topic, our study revealed complex individual patterns of predictors for both aggression and attitudes towards lying. For example, across violent offenders and the general population, higher Neuroticism and lower Agreeableness consistently predicted higher aggression, while higher impression management was associated with lower (reported) aggression. Some population-specific predictors also emerged. Aggression was predicted by a combination of sensitivity to search for excitement (BAS fun seeking), callousness, and self-deceptive enhancement in inmates, and by an increased sensitivity to pursue goals (BAS Drive) in the general population. The next step was to look at the components of aggression (anger, verbal aggression, physical aggression and hostility) separately. Notably, Agreeableness acted as a protective factor against all components of aggression in both populations, while Neuroticism was a risk factor for most components. Interestingly, when looking at actual violent behaviour in prison, only age and gender were identified as predictors, confirming a potential disconnect between (self-reported) trait aggression and situational violence. Regarding attitudes towards lying, across both populations, higher Neuroticism predicted more accepting views, while higher impression management predicted less lenient views. Among violent offenders, higher callousness predicted more positive attitudes towards lying.

The present study significantly contributes to the theoretical and methodological understanding of undesirable behaviours, more specifically aggression, violent behaviours

committed in prison and attitudes towards lying, across diverse populations and settings. Our findings underline the significance of considering a constellation of factors, including personality traits, sensitivity to reward and punishment (BIS/BAS), Callous-Unemotional (CU) traits, and social desirability (impression management and self-deceptive enhancement) factors in explaining aggression and attitudes towards lying. Moreover, the influence of baseline factors such as age and gender could play a role in violent behaviours committed in certain situations. While the current study provides valuable insights, future research should aim to replicate these findings in different populations and further explore the specificity of predictors for various degrees of severity and different offence types. Additionally, extending this work to include observed behaviours and a wider range of situational factors would further enhance our understanding of these complex phenomena. In conclusion, this study highlights the importance of a multifaceted approach in researching undesirable behaviours, emphasising the need for both specificity and generalisability in our understanding of aggression and deception across diverse populations.

4 CHAPTER IV. GENERAL DISCUSSION AND CONCLUSIONS

The overarching aim of the current thesis was to expand the theoretical and empirical understanding of false beliefs and, more particularly, self-deception. More specifically, the main aim was to identify and investigate individual and contextual factors which influence false beliefs and self-deception, and to explore some of the consequences of engaging in self-deception in different settings. For the current thesis, a theoretical model on factors shaping false beliefs and self-deception was proposed, integrating previous models of false beliefs and self-deception with, albeit limited, empirical evidence regarding false beliefs and self-deception (see General Introduction). As false beliefs and self-deception have common underlying elements and characteristics, the proposed integrative model brings together false beliefs research, such as models focusing mostly on errors in belief and recollection, and models which focus on self-deception, predominantly defined as a case of motivated false beliefs.

Throughout the studies included in this thesis, we investigated individual (moral self-concept maintenance as an internal motivation for self-deception in Study 2) and contextual (the role of event plausibility and frequency on false beliefs and memories in Study 3) factors shaping false beliefs and self-deception formation, as well as the consequences and implications of self-deception on memory (self-serving memory bias and memory errors, Study 2) and for self-report measures (aggression and attitudes towards lying as a proxy for deception) in different populations (the general population and violent offenders, Study 4).

In Study 1, we found that self-deception is highly prevalent and frequently used in everyday situations, motivated by predominantly internal motivations (e.g., feeling better), but also external (e.g., avoiding punishments). Additionally, examples of self-deception were focused on exaggerating as well as denying or rejecting information, and the impact of self-deception was more negative than positive in the short term, with the effects diminishing in the long term. In Study 2, we successfully elicited self-deception using a typical paradigm used in this line of research (the forward-thinking paradigm, Chance et al., 2011) and showed that higher self-deceivers attributed their performance to their own abilities when being exposed to a possibility to cheat (access to answers). Additionally, we showed that when participants were exposed to the possibility of cheating by having access to answers compared with no access to answers, they made more memory errors one week later when asked to remember their past choices. And while memory errors were not associated with self-deception, their past performance estimations were, as participants consistently inflated their estimations when

access to answers was provided. Lastly, we showed that moral self-concept was positively related to self-deception; however, this was found only for dispositional self-deception. In Study 3, we successfully elicited false beliefs (and false memories) for past autobiographical events, using a blind implantation paradigm (Otgaar et al., 2022b). Specifically, we elicited false beliefs and false memories for a high plausible ("lost a toy") and low plausible ("almost drowned in the ocean") event, suggested as having allegedly happened once or repeatedly during childhood. And we showed that plausibility and frequency affect the formation of false beliefs, with higher false belief ratings for high than low plausible events, suggested as having occurred one time. In Study 4, we found that there were no differences in self-deception as a more stable measure between the general population and a sample of violent offenders, and that self-deception was negatively associated with measures of self-reported aggression and attitudes towards lying; however, these associations varied depending on context.

4.1 Theoretical and Empirical Contributions

4.1.1. Types of False Beliefs and Self-Deception

Extending the existing theoretical and empirical findings, throughout the studies in this thesis, we successfully elicited and assessed the above-mentioned various types of self-deception and false beliefs. Specifically, we first demonstrated the high prevalence and situational nature of self-deception among university students, providing concrete examples and distinguishing between situational and dispositional self-deception (Study 1). Next, we showed that self-deception and self-serving memory biases can be elicited using validated paradigms and showed that situational and dispositional self-deception measures are associated in a more neutral context only (Study 2). We showed that dispositional self-deception was higher than normative data in a sample from the general population and a sample of violent offenders (Study 4). Finally, we successfully elicited false autobiographical beliefs and memories through a blind implantation paradigm (Study 3). These various types of false beliefs address the core aims of the thesis and provide a structured foundation for the detailed explanations and discussions of this chapter.

4.1.2. Individual and Contextual Factors of False Beliefs and Self-Deception

One of the main theoretical aims of this thesis was to identify individual and contextual factors shaping false beliefs and self-deception. We found that contextual factors such as event plausibility and frequency interact in shaping false beliefs (Study 3), that self-deception is motivated by predominantly internal factors (e.g., feeling better), but also external (e.g., avoiding punishments) (Study 1) and as a dispositional trait, it is associated with moral self-

construct (Study 2). Moreover, dispositional self-deception was not found to be different across settings (e.g., in the general population and violent offenders, Study 4).

4.1.3. Consequences of False Beliefs and Self-Deception

The final theoretical aim of this current thesis was to explore negative (and potentially positive consequences of self-deception. We found that self-deception had negative and positive consequences in everyday situations, which diminished in the long term (Study 1), and it is associated with a self-serving memory bias, but not with memory errors (Study 2). Additionally, self-deception was associated negatively with measures of aggression and attitudes towards lying, depending on the setting (Study 4). Implications for treating self-deception as a more dispositional measure rather than response bias are discussed, as well as how these findings reflect on the integrative model proposed.

For an overview of the aims and main theoretical and empirical contributions of the current thesis, see Table 4.1.1. below. Furthermore, in the current thesis, based on the theoretical and empirical literature on false beliefs and self-deception, an integrative model of individual and contextual factors shaping false beliefs and self-deception, as well as negative and positive consequences of false beliefs and self-deception, was proposed. Figure 4.1.1. shows a graphical representation of the studies included in this thesis and the factors and consequences addressed.

Table 4.1.1.

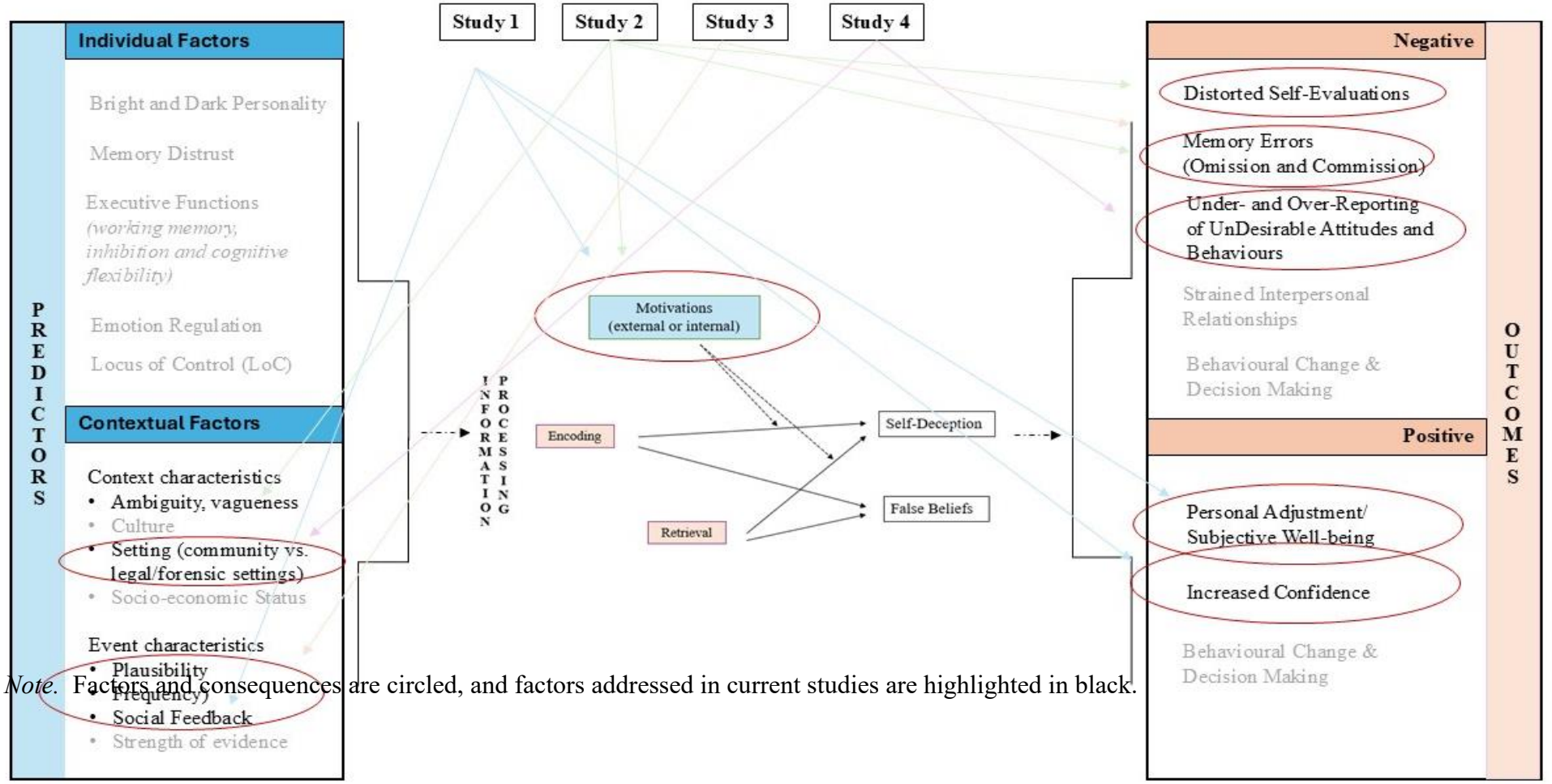
The Main Theoretical and Empirical Contributions of the Current Thesis

Study	Type of false belief	Testing paradigm	Sample	N	Aim	Main Theoretical and Empirical Contributions
1	Self-deception (past event, experiences) and dispositional self-deception	Mixed-methods survey	University students	228	Measure the prevalence and frequency of self-deception in everyday situations, gather examples and explore its elements and characteristics identified in the literature	Showed that self-deception is highly prevalent (91.3% of participants reported engaging in self-deception throughout their life) and reported as frequent, mostly situational than trait, however, negatively associated with a validated measure of self-deception (Paulhus Deception Scales, Paulhus, 1998). Gathered 107 examples of self-deception from university students. Showed that examples and reports included cases of denying or rejecting information, and exaggerating or enhancing information. Identified motivations for engaging in self-deception: intrapersonal motivations were reported as predominant motives for engaging in self-deception; however, interpersonal motivations were also reported. Motivations identified from examples were mostly oriented towards avoiding punishment or negative consequences, followed by feeling better, while self-reported motivations were predominantly feeling better about oneself. Identified examples and frequency of positive and negative consequences of self-deception in everyday life. Identified elements and characteristics relevant to the construct of self-deception (e.g., strategies for maintenance, retraction reasons, etc.).
2	Situational and dispositional self-deception	Forward-thinking paradigm (Chance et al, 2011)	University students	137	Investigate the role of moral self-concept maintenance in self-deception and explore the effects of self-deception on memory (estimations and errors)	Successfully elicited self-deception using a modified version of the forward-thinking paradigm (Chance et al., 2011); additionally showed that participants exposed to a cheating opportunity (with access to correct answers to a test) attribute more their performance to their intellectual abilities, than participants in a more neutral and honest context (without access to answers and possibility to cheat). Showed that moral self-concept was relatively stable (one week difference) and that situational self-deception was not associated with moral self-concept even in a cheating possibility situation;

Study	Type of false belief	Testing paradigm	Sample	N	Aim	Main Theoretical and Empirical Contributions
						Showed that a more positive moral self-concept was positively associated with higher dispositional self-deception. Showed that self-serving memory bias for past abilities was successfully elicited and associated with estimations across time (present, future). Higher present estimations in a cheating possibility group were consistently higher across the three measures than in a group with no cheating possibility. Showed that participants with access to answers in a cheating possibility context had more memory errors in a recognition test; however, this was not associated with self-deception.
3	Autobiographical false beliefs and memories	Blind Implantation paradigm (Otgaar et al., 2022b)	University students	103	Assess the role of plausibility and frequency and their interaction in eliciting false beliefs and memories	Successfully elicited false beliefs and memories for autobiographical events, high plausible ("lost a toy") and low plausible ("almost drowned in the ocean"), which were suggested as having allegedly happened once or repeatedly during childhood (using the blind implantation paradigm, Otgaar et al., 2022). Showed that event plausibility and frequency affect the formation of false beliefs and memories, with high plausibility leading to higher false beliefs than low plausibility when an event was suggested as having allegedly occurred once during childhood.
4	Dispositional self-deception (Self-deceptive enhancement)	Self-deception as a component of social desirability (Paulhus & Reid, 1991)	General Population & Violent Offenders	238 202	Explore group differences in dispositional self-deception in two samples (general population and violent offenders) and explore the added predictive value for measures of aggression and attitudes towards lying	Showed that self-deceptive enhancement is not different in a sample from the general population and individuals serving a sentence for a violent offence Showed that self-deceptive enhancement is associated negatively with aggression in both samples (respectively with anger, hostility and physical aggression) and negatively with attitudes towards lying in violent offenders Showed that self-deceptive enhancement contributed to predicting aggression in the general population and a sample of violent offenders, but not in predicting attitudes towards lying.

Figure 4.1.1.

Overview of Factors and Consequences Investigated in the Current Thesis



Note. Factors and consequences are circled, and factors addressed in current studies are highlighted in black.

4.2 Methodological and Practical Contributions

While studying self-deception comes as a challenge due to ongoing theoretical debates and significant gaps in empirical research, this carries over in all aspects of research, such as methods of exploring and identifying self-deception, providing its own set of challenges. One of the main challenges is the distinction between self-deception and other deception. Throughout the studies in this thesis, we addressed (or attempted to) some of these challenges, as well as offered further recommendations on the best ways to address self-deception from both a more fundamental point of view as well as an applied perspective.

Inspired by the grounded theory approach (Strauss & Corbin, 1994) in which a theory derives from systematically collected and analysed data, we developed a mixed-methods survey which contained open-ended, as well as multiple-choice questions on experiences of self-deception (Study 1). $N = 228$ university students answered our survey, and while the number of participants who responded to each question varied, we succeeded in gathering more than 100 examples of self-deception in everyday situations, which were further analysed and compared with the theories and perspectives presented in the literature. We also extended the forward-thinking paradigm, which is being increasingly used in the literature (Chance et al., 2011, 2015; Jian et al., 2019; Mei et al., 2023) (Study 2). Previous studies proposed that self-deceivers would attribute their higher scores to their abilities rather than access to answers, thus really deceiving themselves (Chance et al., 2011, 2015); however, this was not directly tested. In our study, we included a measure of task performance attribution (Robins & Beer, 2001) and showed that participants with access to answers and higher positive inflated predictions were more likely to attribute their performance to their intellectual abilities than other more externally oriented options (e.g., the option "other" which was intentionally added to be used as a way to express possible cheating). While this method does not directly answer the self vs other deception, it does offer more support for the internalised false belief perspective. We recommend that future studies include ways in which the ground truth is actually measured (e.g., Riesthuis et al., 2021).

On the other hand, we discussed in Study 2 that in the absence of actual overt cheating, these differences in situational self-deception could be attributed to a miscalibration of abilities. The forward-thinking paradigm is superior to other measures of laboratory-elicited self-deception (Chance et al., 2011); however, in light of these findings, we recommend that, if the aim is to test situational self-deception rather than self-serving tendencies, one might adapt paradigms from interpersonal deception (e.g., focused on fabrication, feigning, false denials, Battista & Otgaar, 2022; Otgaar et al., 2023), which measure deception regarding events or

event details that are also more similar to real life examples of self-deception (see Study 1). Lastly, as our focus was also on autobiographical false beliefs, we used and extended a novel paradigm developed to address limitations in previous paradigms studying false beliefs and false memories, such as lengthy interview training or the need for information regarding the used events from others (e.g., parents)(Otgaar et al., 2022b). We found that plausibility played a role in the formation of false beliefs about the event suggested to have happened once. These findings emphasise the importance of the plausibility of events (Scoboria et al., 2004) and interaction effects in contextual factors when designing experiments using this paradigm. Furthermore, we were also interested in the facet of self-deception measured as response bias or trait and dispositional, and how this plays a role in other measures (Study 4).

Overall, for the current thesis, using diverse methods, we gathered data using experimental studies (Studies 2 and 3), as well as in more ecological contexts (Studies 1 and 4). Additionally, we collected data on broader examples and characteristics of self-deception and dispositional self-deception (Study 1), situational self-deception and dispositional self-deception in an experimental context of cheating (Study 2), false autobiographical beliefs in a laboratory setting (Study 3) and dispositional self-deception across different settings, specifically between the general population and violent offenders (Study 4). Gathering data from a variety of perspectives and methods, while not leading to a main conclusion, portrays a broader perspective of the constructs studied.

The focus of the current thesis was to contribute to the field of legal psychology. Namely, we were also interested in the practical and applied aspects of false beliefs and especially self-deception in this field. In the first study included in this thesis, we showed that self-deception was highly prevalent and frequent in day-to-day situations, and some of those situations were contexts in which participants lied to others or feigned being ill (approximately 5% of reasons reported for self-deception). The percentages associated with more legally relevant behaviours could be interpreted as low (however, see Wade et al., 2024 for a critical discussion on addressing percentages in false memories), and the lies were not necessarily severe (“That something was my fault, but I said it was someone else's fault.”). However, in more high-stakes contexts such as the legal field, lying and feigning could have severe implications, such as miscarriages of justice (Brainerd & Reyna, 2005; Merckelbach & Merten, 2012). We showed that examples of self-deception also included examples of falsely denying or rejecting relevant information, which is relevant in a legal setting where false denials are encountered (e.g., witnesses, victims or suspects falsely deny events; Bücken et al., 2023). Our two experimental

studies were rather fundamental; however, still relevant due to their focus on factors shaping false beliefs and self-deception. Moral self-concept was measured in Study 2, and we showed that it was not associated with situational self-deception, but it was associated with dispositional self-deception for both groups. Understanding the role of moral self-concept in self-deception is crucial, as it has been repeatedly assumed that individuals engage in self-deception to maintain their positive image (Baumeister, 1996). While we expected self-deception as a motivated self-generated lie to be associated with memory errors, we did not find this effect. In Study 3, we showed that false beliefs and false memories could be implanted for more or less plausible, allegedly having occurred once or multiple times, events. Furthermore, for events falsely suggested as having occurred once, false beliefs were higher for the high plausible than the low plausible, and this effect was absent when the events were suggested as having occurred multiple times. Lastly, in Study 4, we included a sample relevant to legal psychology, as well as self-reported measures which are relevant for risk assessment and criminal recidivism and untimely legal decision-making, such as aggression and attitudes towards lying (Hurezan et al., 2024; Matlasz et al., 2020). Contrary to what is expected, we found no group differences between the two populations, the general population and violent offenders, in dispositional self-deception.

We showed, in Study 4, that self-deceptive enhancement was negatively associated with aggression (and its components: anger, hostility, and physical aggression) and with attitudes towards lying in the general population. Taken together, this suggests that although self-deceptive enhancement may not be higher in offender samples, contextual factors might influence its role across various measures. Additionally, self-deceptive enhancement significantly predicted self-reported aggression only among violent offenders, not in the general population. Thus, violent offenders with a more positive self-view reported less aggression than those with a less favourable view, an effect observed specifically in anger and physical aggression, but not hostility or verbal aggression. Moreover, in terms of attitudes towards lying, self-deceptive enhancement did not account for any variability (notably, impression management did account for some of these effects, see also Mills & Kroner, 2005). In summary, while self-deceptive enhancement might play a role in specific outcomes relevant to the legal field, our findings emphasise the importance of assessing social desirability as two factors, and the importance of contextual factors (see also Hildebran et al., 2018; Tan & Grace, 2008).

When discussing the **practical relevance** of investigating factors shaping false beliefs and self-deception and the consequences of false beliefs and especially of self-deception, a

central issue in psychology, and particularly legal psychology, is whether such findings are also practically relevant (see Otgaar et al., 2022c; Riesthuis et al., 2025). In other words, are the effects obtained large enough to have clear implications for the legal field? To address this, Table 4.1.2 provides an overview of the main contributions of each study and effect sizes obtained in the studies included in this thesis. Regarding self-deception, we primarily measured estimations of abilities (situational self-deception, Study 2) or self-reports of desirable qualities (dispositional self-deception, Study 2 and Study 4). These types of self-reports of how well one will do or did in a relevant task could be of particular importance, for example, when interviewing experts in court, as courts consider expert confidence when evaluating testimony quality (Neal et al., 2018). If we apply similar reasoning with the one detail difference in memory research (Otgaar et al., 2022c), one point difference in one's estimation of their ability (e.g., how many identifications I did which were a match) might make the difference in crucial decisions (for example, for an expert to consider their decision is ready for court) as the estimation difference we found was of 1.89, thus showing a practically relevant result. Furthermore, it is important to consider the practical significance of the associations and predictive value between measures of self-deception (as a response style or disposition) and self-reported aggression or attitudes toward lying. Specifically, what constitutes a minimal change in aggression or attitudes toward lying that would be considered practically significant (i.e., the minimal clinically important difference; Anvari & Lakens, 2020)? The significant associations found in our studies indicate that the probability of these results occurring by chance ranged from 68% to 72%, based on the Dunlap Common Language Effect Size (CLES), which expresses effect sizes in intuitive probability terms (Riesthuis et al., 2025). There is a 68% to 72% chance that a random individual from the general population who scores high on measures of self-deception will also have a chance of a higher score on measures of moral self-concept, and a lower score on measures of aggression. Similarly, there is a 68% to 72% chance that a random person serving a sentence for committing a violent offence has a higher chance of scoring lower on measures of aggression and attitudes towards lying than a randomly selected individual who scores low on a measure of self-deception. Since these probabilities are meaningfully higher than chance, it is relevant to consider self-deceptive enhancement when using these measures in practice.

Table 4.1.2.

The Methodological and Practical Contributions of the Current Thesis

Study	Aim	Main Methodological and Practical Contributions	Unstandardised Effect Size	Standardised Effect Size
1	Develop a novel method for investigating and gathering examples of self-deception	Showed that using a mixed-method survey, self-deception can be studied retrospectively.	200 out of 228	91.3% (200/228)
		Gathered examples of self-deception in a more applied, ecological context (day-to-day situations and contexts).	NA	NA
2	Extend a method of situational self-deception (forward-thinking paradigm, Chace et al., 2011) to include memory effects	Showed that situational self-deception can be elicited using the forward-thinking paradigm (Chance et al., 2011), also online, even in the absence of overt cheating	Mean difference (test with 20 answers, range 1-20)	Cohen's *d = 0.94*
		Showed that self-serving memory bias can be elicited for past performance estimation using the forward-thinking paradigm	M_{diff} = 1.89 M_{diff} = 2.66	Cohen's *d = 0.80*
		Showed that being exposed to a cheating opportunity (access to correct answers for a test) leads to more memory errors (recognition of answers one week later).		
		Showed that measures of moral self-concept can be impacted by self-deception, as dispositional self-deception was associated with moral self-concept	M_{diff} = 2.16	Cohen's *d* = 0.71
			Unst. Reg. Coef. 0.12 0.07	Pearson*'s r* *r* = .36 *r* = .33
3	Extend a novel paradigm (blind implantation, Otgaar et al., 2022b) of studying false beliefs and false memories to include event plausibility	Showed that false beliefs for high and low plausible events, as well as suggested allegedly happened one or multiple times during one's childhood, can be formed using this paradigm	13 out of 25 10 out of 26 3 out of 30 5 out of 22	52% (n = 13) 38.5% (n = 10) 10% (n = 3) 22.7% (n = 5)
		Showed that event plausibility and frequency interact in forming false beliefs and memories, with a main effect of interaction for simple effects of plausibility; the single high plausible event had higher false belief ratings than the low plausibility event.	31 out of 103 Mean difference = 2.49 (8-point Likert scale)	30.1% (n= 31) η_p^2 = 0.05 Cohen's d = 0.99

Study	Aim	Main Methodological and Practical Contributions	Unstandardised Effect Size	Standardised Effect Size
4	Explore if self-deceptive enhancement adds variability to measures of aggression and attitudes towards lying in the general population and violent offenders using self-deceptive enhancement as a component of social desirability	Showed no statistically significant group differences between the general population and violent offenders in self-deceptive enhancement.	Mean difference = 1.85 (20 items, each rated 1–5, total score range 20–100)	Cohen's *d* = - .23 (Bonferonni corrections for multiple comparisons, sign at p < .0025)
		Showed that self-deceptive enhancement is associated negatively with aggression in both samples (respectively with anger, hostility and physical aggression) and negatively with attitudes towards lying in violent offenders	Unst regr. Coeff. - 0.60, - 0.27, - 0.33, - 0.18, - 0.33, - 0.85, - 0.33, - 0.22, - 0.36	Pearson's *r* -.35, -.39, -.20, -.36;
		Showed that self-deceptive enhancement contributed to predicting aggression in the general population and a sample of violent offenders, but not in predicting attitudes towards lying.	*B* = - 0.30 *B* = - 0.11 *B* = - 0.18	-.33, -.33, -.35, -..24, -.28 *β* = -.13 *β* = -.16 *β* = -.19

Note. When multiple coefficients are included, they are presented in the order they were listed in column three (main contributions). Coefficients have been calculated using the following effect size calculator: https://paulriesthuis.shinyapps.io/interpreteffect/. Means, standard deviations and confidence intervals are reported in each study.

4.3 Practical Implications, Limitations and Conclusions

The current thesis's findings have practically relevant implications for understanding (motivated) false beliefs in the past, present and future in different contexts with direct applications to the legal field. In legal procedures, especially in the courtroom, a major weight is given to eyewitness testimony, and while highly accurate, memory and our belief system are reconstructive processes (Loftus, 2005; Greve & Wentura, 2010). Furthermore, a distinction between belief and recollection (Scoboria et al. 2004) could be highly essential, as some witnesses, for example, might consider that they are reporting their memories of past events, when in fact they might be reporting their beliefs in the occurrence of events. Experts, such as forensic experts handling evidence, could judge their abilities in identifying, for example, suspects' fingerprints as better than in reality, and this might impact court proceedings (Neal et al., 2018). Additionally, in legal contexts, suspects might underreport their behaviours and attitudes when completing measures of self-report (Bianchini et al., 2001; Boskovic et al., 2024; Hildebrand et al., 2018). Understanding how false beliefs and self-deception develop, and how they affect everyday functioning and outcomes relevant to the legal field, such as memory accuracy and the reliability of self-report measures used in legal decision-making, is crucial. Such insights can lead to practical recommendations for preventing, reducing, and managing false beliefs and self-deception both in daily life and in legal contexts, where their impact can be especially serious. However, before research can be used for making recommendations, for example, in expert witness reports, certain requirements have been proposed (Chin & Neal, 2023; Otgaar et al., 2022).

Otgaar and colleagues (2022c) proposed requirements for the generalizability, replicability and practical relevance of research before it can be used, especially in the legal arena. While the findings presented in the current thesis have met some of the requirements for practical relevance, as shown in the previous section, the other two criteria are yet to be met. More specifically, the use of mostly university samples in the more fundamental studies (Study 2 and 3) and the specificity of the contexts for the more applied studies (university students, thus academic contexts offered mostly for self-deception examples, Study 1; specific general population and violent offenders, samples, Study 4) decrease the confidence in their generalizability. Further research is needed to establish whether these findings can be replicated in more diverse and ecologically valid samples and contexts, and until such evidence is available, caution is warranted in extending the current findings beyond the studied samples and contexts.

Although the thesis contributes to the theoretical and empirical literature on false beliefs, particularly regarding self-deception, it is not without limitations. The first limitation relates to one of the main constructs studied through the thesis, namely self-deception. From its definition (e.g., still debated if it is really a motivated type of false belief and which motivations and how these work; Mele, 2000; von Hippel & Trivers, 2011; Moldoveanu & Visu-Petra, 2025), to operationalisation (e.g., how can the awareness of the truth be proven in experimental research, see Study 2). Additionally, some of the findings of this thesis showed that elements or types of self-deception (e.g., rejecting, denying information, see Study 1) might be overlooked in current empirical studies. Future studies should focus on reaching a consensus amongst researchers and experts on the definition and elements of self-deception, further clarify and expand its operationalisation and measurement, and identify boundary conditions (e.g., the limits of when and for whom these phenomena are most likely to occur (Busse et al., 2016).

For the current thesis, we proposed an integrative model of (motivated) false beliefs, which included individual and contextual factors shaping false beliefs and self-deception, as well as several consequences on day-to-day functioning and for more applied settings, such as the legal setting. However, we only addressed some of these factors and consequences. Future studies should systematically test motivational factors and further clarify the conceptual boundaries between constructs. In the studies included in this thesis, various methodologies were used (e.g., survey, experimental paradigms, self-report data) and several aspects of false beliefs and self-deception were investigated (e.g., factors, consequences, etc). This diversity in aspects and methods led to a broader understanding of the constructs investigated. The downside of this is that it is difficult to present a main conclusion alongside practical recommendations. Furthermore, the main samples involved in this dissertation were primarily university students, except for Study 4, in which a community sample and a sample of violent offenders were included. Future studies should also include different samples for a better generalizability of results.

While the aims of this thesis were not focused on cultural aspects of false beliefs and false memories, throughout the studies, our samples were culturally diverse, with samples from WEIRD (Western, Educated, Industrialised, Rich, Democratic) countries (e.g., Belgium, the Netherlands), as well as non-WEIRD populations (e.g., Romania, Iran). While we cannot make conclusions based on culture due to not being directly investigated in our studies, the multi-cultural aspect enriches this thesis and could be informative for studies which focus on cultural

aspects as well. Future studies could target several aspects addressed in this thesis and systematically investigate the role that culture, as a proposed individual factor, plays.

The current thesis proposed and empirically tested across four studies an integrative theoretical model of factors shaping false beliefs and self-deception, bridging previous models of autobiographical belief and motivated false beliefs and existing empirical research. We found that false beliefs were successfully elicited for low, high, suggested as happened once or repeated events, and plausibility played a role when an event was suggested as having allegedly occurred once; self-deception, as exaggerating, or denying information, was highly prevalent in day-to-day situations and was influenced by internal (feeling good) and external motivations (e.g., avoiding punishments, lying to others). Self-deception was not associated with more memory errors in a cheating context; however, it was associated with a self-serving memory bias, where participants inflated their performance estimates. Additionally, self-deception was positively associated with self-reported moral self-concept and negatively associated with self-reported aggression and attitudes towards lying. However, these associations were context-dependent (e.g., varied between context and setting). Contrary to expectations, dispositional self-deception did not differ between the general population and a sample of violent offenders.

Taken together, these results address several gaps in the theoretical and empirical literature and have practical relevance for several areas of the legal field, such as evaluating the validity of testimony and self-report measures. However, the findings presented should be interpreted with caution, and further research should be conducted before being used confidently in courtroom settings. Nonetheless, the present thesis extended theoretical, empirical, and methodological aspects in studying false beliefs and self-deception and brought the phenomenon of self-deception more centrally to the arena of legal psychology research.

REFERENCES

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20,* 1–48. https://doi.org/10.1080/10463280802613866

Althubaiti, A. (2023). Sample size determination: a practical guide for health researchers. *Journal Of General And Family Medicine, 24*(2), 72-78. https://doi.org/10.1002/jgf2.600

Anagnostaki, L., Wright, M. J., & Papathanasiou, A. (2013). Secrets and disclosures: How young children handle secrets. The Journal of Genetic Psychology, 174(3), 316–334. https://doi.org/10.1080/00221325.2012.672350

Anderson, C. A., & Bushman, B. J. (2002). The General Aggression Model: Theoretical extensions to violence. *Psychology of Violence, 13(4*), 245-258. https://doi.org/10.1037/a0023842

Anvari, F., Kievit, R., Lakens, D., Pennington, C. R., Przybylski, A. K., Tiokhin, L., Wiernik, B. M., & Orben, A. (2023). Not All Effects Are Indispensable: Psychological Science Requires Verifiable Lines of Reasoning for Whether an Effect Matters. *Perspectives on Psychological Science, 18*(2), 503-507. https://doi.org/10.1177/17456916221091565

Aquino K., Reed I. I. (2002). The self-importance of moral identity. Journal of Personality and Social Psychology, 83, 1423–1440

Arndt, J. E., Hoglund, W. L., & Fujiwara, E. (2013). Desirable responding mediates the relationship between emotion regulation and anxiety. *Personality and Individual Differences*, *55*(2), 147-151. https://doi.org/10.1016/j.paid.2013.02.022

Bachkirova, T. (2016). A new perspective on self-deception for applied purposes. *New Ideas in Psychology*, *43*, 1-9. https://doi.org/10.1016/j.newideapsych.2016.02.004

Balcetis, E. (2007). Where the motivation resides and self-deception hides: how motivated cognition accomplishes self-deception. Social and Personality Psychology Compass, 2(1), 361-381. https://doi.org/10.1111/j.1751-9004.2007.00042.x

Barry, C. T., Lui, J. H. L., & Anderson, A. C. (2016). Adolescent Narcissism, Aggression, and Prosocial Behavior: The Relevance of Socially Desirable Responding. *Journal of Personality Assessment*, *99*(1), 46–55. https://doi.org/10.1080/00223891.2016.1193812

Battista, F., & Otgaar, H. (2022). Research on the effects of lying on memory: a scientometric analysis and a call for new studies. *Frontiers in Psychology*, *13*, 837265. https://doi.org/10.3389/fpsyg.2022.837265

Battista, F., Curci, A., Mangiulli, I., & Otgaar, H. (2021). What can we remember after complex denials? The impact of different false denials on memory. *Psychology, Crime and Law*. https://doi.org/10.1080/1068316X.2020.1865956

Battista, F., Mangiulli, I., Herter, J., Curci, A., & Otgaar, H. (2020). The effects of repeated denials and fabrication on memory. *Journal of Cognitive Psychology, 32*(4), 369–381. https://doi.org/10.1080/20445911.2020.1767626

Battista, F., Mangiulli, I., Patihis, L., Dodier, O., Curci, A., Lanciano, T., & Otgaar, H. (2023). A scientometric and descriptive review on the debate about repressed memories and traumatic forgetting. *Journal of anxiety disorders*, *97*, 102733. https://doi.org/10.1016/j.janxdis.2023.102733

Battista, F., Otgaar, H., Lanciano, T., & Curci, A. (2020). Individual differences impact memory for a crime: A study on executive functions resources. *Consciousness and Cognition*, *84*, 103000. https://doi.org/10.1016/j.concog.2020.103000

Baumeister, R. F. (1993). Lying to yourself: The enigma of self-deception. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 166-183). Guilford Press.

Baumeister, R. F. (1996). Self-regulation and ego threat: Motivated cognition, self-deception, and destructive goal setting. In P. M. Gollwitzer & J. A. Bargh (Eds.), The psychology of action: Linking cognition and motivation to behavior (pp. 27–47). The Guilford Press.

Baumeister, R. F., & Cairns, K. J. (1992). Repression and self-presentation: When audiences interfere with self-deceptive strategies. Journal of Personality and Social Psychology, 62(5), 851–862. https://doi.org/10.1037/0022-3514.62.5.851

Bays, R. B., Zabrucky, K. M., & Gagne, P. (2012). When plausibility manipulations work: an examination of their role in the development of false beliefs and memories. *Memory* 20(6), 638–644. https://doi.org/10.1080/09658211.2012.692797

Ben-Porath, Y. S., & Tellegen, A. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) scales in mental health, forensic, and nonclinical settings: An introduction. *Journal of Personality Assessment, 90*(2), 119–121. https://doi.org/10.1080/00223890701845120

Bernstein, D. M., Godfrey, R. D., & Loftus, E. F. (2009). False memories: The role of plausibility and autobiographical belief. In K. D. Markman, W. M. P. Klein, & J. A. Suhr (Eds.), *Handbook of imagination and mental simulation* (pp. 89–102). Psychology Press.

Bernstein, D. M., Scoboria, A., & Arnold, R. (2015). The consequences of suggesting false childhood food events. *Acta Psychologica*, *156*, 1-7. https://doi.org/10.1016/j.actpsy.2015.01.001

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom Validity Testing: A Critical Review. *The Clinical Neuropsychologist*, *15*(1), 19–45. https://doi.org/10.1076/clin.15.1.19.1907

Białecka-Pikul, M., Białek, A., Stępień-Nycz, M., Talwar, V., & Bosacki, S. (2022). Peeking and lying in the temptation resistance paradigm in 2.5-year-olds: The role of inhibitory control. *PLOS ONE, 17*(12), e0278099. https://doi.org/10.1371/journal.pone.0278099

Blandón-Gitlin, I., & Gerkens, D. (2010). The effects of photographs and event plausibility in creating false beliefs. *Acta Psychologica, 135*(3), 330-334. https://doi.org/10.1016/j.actpsy.2010.08.008

Brainerd, C., Holliday, R., Reyna, V., Yang, Y., & Toglia, M. (2010). Developmental reversals in false memory: Effects of emotional valence and arousal. *Journal of Experimental Child Psychology, 107*(2), 137-154. https://doi.org/10.1016/j.jecp.2010.04.013

Boskovic, I., Giromini, L., Katsouri, A., Tsvetanova, E., Fonse, J., & Merckelbach, H. (2024). The spectrum of response bias in trauma reports: overreporting, underreporting, and mixed presentation. *Psychological Injury and Law, 17,* 117–128. https://doi.org/10.1007/s12207-024-09503-w

Bücken, C. A., Otgaar, H., London, K., Riesthuis, P., Battista, F., & Mangiulli, I. (2023). 'Nothing happened': Legal implications of false denials among abused children. Child Abuse Review, 32(2), e2791. https://doi.org/10.1002/car.2791

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452-459. https://doi.org/10.1037/0022-3514.63.3.452

Busse, C., Kach, A. P., & Wagner, S. M. (2016). Boundary Conditions: What They Are, How to Explore Them, Why We Need Them, and When to Consider Them. *Organizational Research Methods*, *20*(4), 574-609. https://doi.org/10.1177/1094428116641191

Buta, M., Visu-Petra, G., Opre, A., Koller, S. & Visu-Petra., L. (2021): Individual Predictors Of Lie Acceptability Across Development. *European Journal of Developmental Psychology*, *19*(5), 696–710. https://doi.org/10.1080/17405629.2021.1947234

Butterworth, J., Trivers, R., & Von Hippel, W. (2022). The better to fool you with: Deception and self-deception. *Current Opinion in Psychology*, *47*, 101385. https://doi.org/10.1016/j.copsyc.2022.101385

Calado, B., Luke, T. J., Connolly, D. A., Landström, S., & Otgaar, H. (2021). Implanting false autobiographical memories for repeated events. *Memory, 29*(10), 1320–1341. https://doi.org/10.1080/09658211.2021.1981944

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology, 67(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Chance, Z., & Norton, M. I. (2015). The what and why of self-deception. *Current Opinion in Psychology*, *6*, 104-107. https://doi.org/10.1016/j.copsyc.2015.07.008

Chance, Z., Gino, F., Norton, M. I., & Ariely, D. (2015). The slow decay and quick revival of self-deception. Frontiers in Psychology, 6:1075. https://doi.org/10.3389/fpsyg.2015.01075

Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. Proceedings of the National Academy of Sciences, 108(3), 15655-15659. https://doi.org/10.1073/pnas.1010658108

Clark A, Nash RA, Fincham G, Mazzoni G (2012) Creating non-believed memories for recent autobiographical events. *PLoS ONE 7*(3): e32998. https://doi.org/10.1371/journal.pone.0032998

Cochran, K. J., Greenspan, R. L., Bogart, D. F., & Loftus, E. F. (2016). Memory blindness: Altered memory reports lead to distortion in eyewitness memory. *Memory & Cognition, 44*(5), 717–726. https://doi.org/10.3758/s13421-016-0594-y

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Academic Press.* https://doi.org/10.4324/9780203771587

Costa, P. T., Jr., & McCrae, R. R. (2010). Bridging the gap with the five-factor model. *Personality Disorders: Theory, Research, and Treatment, 1*(2), 127–130. https://doi.org/10.1037/a0020264

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. https://doi.org/10.1037/h0047358

D'Argembeau, A., & Van der Linden, M. (2008). Remembering pride and shame: Self-enhancement and the phenomenology of autobiographical memory. Memory, 16(5), 538–547. https://doi.org/10.1080/09658210802010463

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & 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

Dodier, O., Otgaar, H., & Mangiulli, I. (2024). Beyond repressed memory: Current alternative solutions to the controversy. *Topics in Cognitive Science, 16*(4), 574–589. https://doi.org/10.1111/tops.12754

Dufner, M., Gebauer, J. E., Sedikides, C., & Denissen, J. J. A. (2019). Self-Enhancement and Psychological Adjustment: A Meta-Analytic Review. *Personality and social psychology review: an official journal of the Society for Personality and Social Psychology, Inc*, *23*(1), 48–72. https://doi.org/10.1177/1088868318756467

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*(293), 52-64. http://www.jstor.org/stable/2282330

Eekhout, I., de Vet, H. C., Twisk, J. W., Brand, J. P., de Boer, M. R., & Heymans, M. W. (2014). Missing Data In A Multi-Item Instrument Were Best Handled By Multiple Imputation At The Item Score Level. *Journal of Clinical Epidemiology, 67*(3), 335–342. https://doi.org/10.1016/j.jclinepi.2013.09.009

Erez, A., Johnson, D. E., & Judge, T. A. (1995). Self-deception as a mediator of the relationship between dispositions and subjective well-being. *Personality and Individual Differences*, *19*(5), 597-612. https://doi.org/10.1016/0191-8869(95)00088-N

Falk, O., Wallinius, M., Lundström, S., Frisell, T., Anckarsäter, H., & Kerekes, N. (2014). The 1% of the population accountable for 63% of all violent crime convictions. *Social Psychiatry And Psychiatric Epidemiology*, *49*(4), 559–571. https://doi.org/10.1007/s00127-013-0783-y

Fan, W., Ren, M., Zhang, W., & Zhong, Y. (2022). The impact of feedback on self-deception: Evidence from ERP. *Acta Psychologica Sinica*, *54*(5), 481-496.

Freud, A. (1946). *The ego and the mechanisms of defence.* International Universities Press.

Freud, S. (1960). *The ego and the id.* New York and London.

Freud, S. (1991). The psychopathology of everyday life. (Vol. 5). Penguin Books.

Frick, P. J. (2004). Inventory of Callous–Unemotional Traits. APA PsycTests. https://doi.org/10.1037/t62639-000

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*(2), 208–214. https://doi.org/10.3758/BF03212420

Justice and Behavior, 29(4), 397-426. https://doi.org/10.1177/0093854802029004004

Geraerts, E., Bernstein, D. M., Merckelbach, H., Linders, C., Raymaekers, L., & Loftus, E. F. (2008). Lasting False Beliefs and Their Behavioral Consequences. *Psychological Science*, *19*(8), 749-753. https://doi.org/10.1111/j.1467-9280.2008.02151.x

Geven, L. M., Ben-Shakhar, G., Kindt, M., & Verschuere, B. (2020). Memory-based deception detection: extending the cognitive signature of lying from instructed to self-initiated cheating. Topics in cognitive science, 12(2), 608–631. https://doi.org/10.1111/tops.12353

Gramzow, R. H., & Willard, G. (2006). Exaggerating current and past performance: motivated self-enhancement versus reconstructive memory. *Personality and Social Psychology Bulletin, 32*(8), 1114–1125. https://doi.org/10.1177/0146167206288600

Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology, 9*(1), 103–119. https://doi.org/10.1348/135532504322776889

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35*(7), 603–618. https://doi.org/10.1037/0003-066X.35.7.603

Greve, W., & Wentura, D. (2010). True lies: Self-stabilization without self-deception. Consciousness and Cognition: An International Journal, 19(3), 721–730. https://doi.org/10.1016/j.concog.2010.05.016

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Guenther, C. L., Zhang, Y., & Sedikides, C. (2023). The Authentic Self Is the Self-Enhancing Self: A Self-Enhancement Framework of Authenticity. *Personality and Social Psychology Bulletin*, *50*(8), 1182-1196. https://doi.org/10.1177/01461672231160653

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. Journal of Personality and Social Psychology, 37(2), 147–169. https://doi.org/10.1037/0022-3514.37.2.147

Hampton, R. S., & Varnum, M. E. W. (2018). Do cultures vary in self-enhancement? ERP, behavioural, and self-report evidence. *Social Neuroscience, 13*(5), 566–578. https://doi.org/10.1080/17470919.2017.1361471

Hildebrand, M., Wibbelink, C. J., & Verschuere, B. (2018). Do impression management and self-deception distort self-report measures with content of dynamic risk factors in offender samples? A meta-analytic review. *International Journal of Law and Psychiatry, 58*, 157-170. https://doi.org/10.1016/j.ijlp.2018.02.013

Hirschfeld, R. R., Thomas, C. H., & McNatt, D. B. (2007). Implications of Self-Deception for Self-Reported Intrinsic and Extrinsic Motivational Dispositions and Actual Learning Performance: A Higher Order Structural Model. *Educational and Psychological Measurement, 68*(1), 154-173. https://doi.org/10.1177/0013164406299129

Hornsveld, R. H. J., & Kraaimaat, F. W. (2023). Psychological factors predicting violent prison inmates' anger and aggression. *Journal of Forensic Psychology Research and Practice, 23*(5), 472–488. https://doi.org/10.1080/24732850.2022.2048768

Howe, M. L., & Knott, L. M. (2015). The fallibility of memory in judicial processes: lessons from the past and their modern consequences. *Memory, 23*(5), 633–656. https://doi.org/10.1080/09658211.2015.1010709

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, *60*(2), 229-251. https://doi.org/10.1016/j.jml.2008.10.002

Hurezan, L., Turi, A., Ion, A., & Visu-Petra, L. (2024). Dark and bright personality dimensions as predictors of criminal behavior and recidivism. *Scientific reports, 14*(1), 18565. https://doi.org/10.1038/s41598-024-69288-5

Iliescu, D., & Sîrbu, A. A. (2019). Inventarele NEO: Manual tehnic și interpretativ. [NEO inventories: Technical and interpretative manual]. Sinapsis Publishing Projects.

Jackson C. A. (2009). The Life Events Inventory (LEI). *Occupational medicine (Oxford, England)*, *59*(3), 208. https://doi.org/10.1093/occmed/kqn181

Jian, Z., Zhang, W., Tian, L., Fan, W., & Zhong, Y. (2019). Self-Deception Reduces Cognitive Load: The Role of Involuntary Conscious Memory Impairment. *Frontiers in Psychology*, *10*, 468325. https://doi.org/10.3389/fpsyg.2019.01718

Johnson, E. A. (1995). Self-deceptive coping: Adaptive only in ambiguous contexts. Journal of Personality, 63(4), 759–791. https://doi.org/10.1111/j.1467-6494.1995.tb00316.x

Johnson, E. A., Vincent, N., & Ross, L. (1997). Self-deception versus self-esteem in buffering the negative effects of failure. *Journal of Research in Personality, 31*(3), 385–405. https://doi.org/10.1006/jrpe.1997.2187

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice, 39*(4), 329-337. https://doi.org/10.1016/j.jcrimjus.2011.03.004

Jordan, J., Leliveld, M. C., & Tenbrunsel, A. E. (2015). The moral self-image scale: Measuring and understanding the malleability of the moral self. Frontiers in Psychology, 6, 1878. https://doi.org/10.3389/fpsyg.2015.01878

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist, 52*(3), 221–233. https://doi.org/10.1037/0003-066X.52.3.221

Khalil, E. L. (2017). Making Sense of Self-Deception: Distinguishing Self-Deception from Delusion, Moral Licensing, Cognitive Dissonance and Other Self-Distortions. *Philosophy*, *92*(362), 539–563. https://www.jstor.org/stable/26419333

Khalil, E. L., & Feltovich, N. (2018). Moral licensing, instrumental apology and insincerity aversion: Taking Immanuel Kant to the lab. *PLoS ONE, 13*(11), Article e0206878. https://doi.org/10.1371/journal.pone.0206878

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. Medical teacher, 42(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Kolarcik, P., Madarasova Geckova, A., Reijneveld, S. A., & van Dijk, J. P. (2016). Delinquent and aggressive behavior and social desirability among roma and non-roma adolescents in Slovakia: a cross-sectional study. *Journal Of Interpersonal Violence*, *31*(4), 677–693. https://doi.org/10.1177/0886260514556103

Komasi, S., Saeidi, M., Soroush, A., & Zakiei, A. (2016). The relationship between brain behavioral systems and the characteristics of the five-factor model of personality with aggression among Iranian students. *Journal Of Injury & Violence Research*, *8*(2), 67–74. DOI: 10.5249/jivr.v8i2.696

Krickel, B. (2024). A psychological "how-possibly" model of repression. *Neuropsychoanalysis, 26*(2), 159–174. https://doi.org/10.1080/15294145.2024.2374237

Kunst, M. J. J., & Winkel, F. W. (2015). PTSD symptom reporting and persistent malingering in recipients of state compensation for violent crime victimization: an exploratory study. *The Journal of Forensic Psychiatry & Psychology*, *26*(3), 309–324. https://doi.org/10.1080/14789949.2015.1004636

Lamba, S., & Nityananda, V. (2014). Self-Deceived Individuals Are Better at Deceiving Others. *PLOS ONE*, *9*(8), e104562. https://doi.org/10.1371/journal.pone.0104562

Lampinen, J.M., Neuschatz, J.S., &Payne, D.G. (1998). Memory illusions and consciousness: Examining the phenomenology of true and false memories, Current Psychology. *16*:181-224

Laney, C., Morris, E. K., Bernstein, D. M., Wakefield, B. M., & Loftus, E. F. (2008). Asparagus, a love story: Healthier eating could be just a false memory away. *Memory, 55*(5), 291–298. https://doi.org/10.1027/1618-3169.55.5.291

Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity of scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement, 67*(5), 859–876. https://doi.org/10.1177/0013164406299104

Lauria, F., Preissmann, D., & Clément, F. (2016). Self-deception as affective coping. An empirical perspective on philosophical issues. *Consciousness and cognition*, *41*, 119–134. https://doi.org/10.1016/j.concog.2016.02.001

Levine, E. E., & Lupoli, M. J. (2022). Prosocial lies: Causes and consequences. *Current Opinion In Psychology, 43*, 335–340. https://doi.org/10.1016/j.copsyc.2021.08.006

Lewis, M. (2007). Stepwise versus hierarchical regression: Pros and cons. *Southwest Educational Research Association*. https://files.eric.ed.gov/fulltext/ED534385.pdf

Li, Y., Liu, Z. (2024). The effects of deception on memory: a comparative study of actors and eyewitnesses accounts. BMC Psychol 12, 638. https://doi.org/10.1186/s40359-024-02152-3

Li, C., Otgaar, H., Muris, P., Zhang, Y. & Wang, J. (2024). Inducing emotionally negative nonbelieved memories using negative pictures. *Mem Cogn 52*, 41–56. https://doi.org/10.3758/s13421-023-01441-3

Li, C., Wang, J., & Otgaar, H. (2020). Creating nonbelieved memories for bizarre actions using an imagination inflation procedure. *Applied Cognitive Psychology*, *34*(6), 1277-1286. https://doi.org/10.1002/acp.3722

Liu, J., Ding, W., Deng, L., Tan, M., & Guan, P. (2025). The influence of different negative feedback on the decay of self-deception. *Frontiers in Psychology*, *15*, 1499089. https://doi.org/10.3389/fpsyg.2024.1499089

Liu, J., Zhang, W., Zhan, Y., Song, L., Guan, P., Kang, D., Jian, J., Cai, R., & Li, M. (2019). The effect of negative feedback on positive beliefs in self-deception. *Frontiers in Psychology, 10,* Article 702. https://doi.org/10.3389/fpsyg.2019.00702

Loftus E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory (Cold Spring Harbour, N.Y.)*, *12*(4), 361–366. https://doi.org/10.1101/lm.94705

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory: false memories and allegations of sexual abuse.* St. Martin's Press.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*(12), 720–725. https://doi.org/10.3928/0048-5713-19951201-07

Lu, H. J., & Chang, L. (2011). The association between self-deception and moral self-concept as functions of self-consciousness. *Personality and Individual Differences, 51*(7), 845–849. https://doi.org/10.1016/j.paid.2011.07.014

Lu, H.-J. (2012). Self-deception: Deceiving yourself to better deceive others. *Acta Psychologica Sinica, 44*(9), 1265–1278. https://doi.org/10.3724/SP.J.1041.2012.01265

Magnusson, K. (2023). A Causal Inference Perspective on Therapist Effects. PsyArXiv. https://rpsychologist.com/cohend/

Mathie, N. L., & Wakeling, H. C. (2011). Assessing socially desirable responding and its impact on self-report measures among sexual offenders. *Psychology, Crime & Law*, *17*(3), 215–237. https://doi.org/10.1080/10683160903113681

Matlasz, T. M., Frick, P. J., & Clark, J. E. (2020). Understanding the social relationships of youth with callous-unemotional traits using peer nominations. *Journal of Clinical Child & Adolescent Psychology*, *49*(6), 1-12. https://doi.org/10.1080/15374416.2020.1823847

Matlasz, T. M., Frick, P. J., Robertson, E. L., Ray, J. V., Thornton, L. C., Wall Myers, T. D., Steinberg, L., & Cauffman, E. (2020). Does self-report of aggression after first arrest predict future offending and do the forms and functions of aggression matter? *Psychological Assessment*, *32*(3), 265–276. https://doi.org/10.1037/pas0000783

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*. https://doi.org/10.1509/jmkr.45.6.633

Mazzoni, G. A., Loftus, E. F., & Kirsch, I. (2001). Changing beliefs about implausible autobiographical events: A little plausibility goes a long way. *Journal of Experimental Psychology: Applied, 7(*1), 51-59. https://pubmed.ncbi.nlm.nih.gov/11577619/

Mazzoni, G., & Kirsch, I. (2002). Autobiographical memories and beliefs: A preliminary metacognitive model. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 121–145). Cambridge University Press. https://doi.org/10.1017/CBO9780511489976.007

Mazzoni, G., & Memon, A. (2003). Imagination can create false autobiographical memories. *Psychological Science*, *14*(2), 186-188. https://doi.org/10.1046/j.1432-1327.1999.00020.x

Mazzoni, G., Clark, A., & Nash, R. (2014). Disowned recollections: Denying true experiences undermines belief in occurrence but not judgments of remembering. *Acta Psychologica*, *145*, 139–146. https://doi.org/10.1016/j.actpsy.2013.11.007

Mazzoni, G., Scoboria, A., & Harvey, L. (2010). Nonbelieved memories. *Psychological Science*, *21*(9), 1334-1340. https://doi.org/10.1177/0956797610379865

McEwan, T. E., Davis, M. R., MacKenzie, R., & Mullen, P. E. (2009). The effects of social desirability response bias on STAXI-2 profiles in a clinical forensic sample. British *Journal of Clinical Psychology, 48*(4), 431-436. https://doi.org/10.1348/014466509X454886

McNally, R. J. (2003). Remembering trauma. Belknap Press/Harvard University Press
Mei, D., Ke, Z., Li, Z., Zhang, W., Gao, D., & Yin, L. (2023). Self-deception: Distorted metacognitive process in ambiguous contexts. *Human Brain Mapping*, *44*(3), 948-969. https://doi.org/10.1002/hbm.26116
Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences, 20*(1), 91–136. https://doi.org/10.1017/S0140525X97000034
Mele, A. R. (1998). Noninstrumental Rationalizing. *Pacific Philosophical Quarterly*, *79*(3), 236-250. https://doi.org/10.1111/1468-0114.00060
Mele, A. R. (2000). Self-deception and emotion. *Consciousness & Emotion, 1*(1), 115–137. https://doi.org/10.1075/ce.1.1.07mel
Mele, A. R. (2001). *Self-deception unmasked*. Princeton University Press.
Merckelbach, H., & Merten, T. (2012). A Note on Cognitive Dissonance and Malingering. *The Clinical Neuropsychologist*, *26*(7), 1217–1229. https://doi.org/10.1080/13854046.2012.710252
Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*(2), 122–137. https://doi.org/10.1007/s12207-013-9155-x
Mijović-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*(1538), 227–240. https://doi.org/10.1098/rstb.2009.0218
Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*(2), 213–225. https://doi.org/10.1037/h0076486
Mills, J. F., & Kroner, D. G. (2005). An investigation into the relationship between socially desirable responding and offender self-report. *Psychological Services, 2(1),* 70-80.
Mills, J. F., & Kroner, D. G. (2006). Impression Management and Self-Report Among Violent Offenders. *Journal of Interpersonal Violence*, *21*(2), 178-192. https://doi.org/10.1177/0886260505282288
Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. *Criminal Behaviour and Mental Health, 13*(2), 140-150. https://doi.org/10.1002/cbm.536
Mirandola, C., Lanciano, T., Battista, F., Otgaar, H., & Curci, A. (2023). Psychopathic personality traits are linked to reduced false memories for negative events. *British journal of psychology (London, England : 1953)*, *114*(1), 176–193. https://doi.org/10.1111/bjop.12604
Mirandola, C., Toffalini, E., Ferruzza, E., & Pazzaglia, F. (2020). The power of extraversion? Reduced false memories for positive events. *Personality and Individual Differences*, *159*, 109861. https://doi.org/10.1016/j.paid.2020.109861
Moldoveanu, M.G., & Visu-Petra, L. (2025). Self-deception beyond speculation: A narrative review of the empirical research on motivated false beliefs. *Studia Universitatis Babes-Bolyai, Psychologia-Paedagogia*
Moldoveanu, M.G., Visu-Petra, L., Otgaar, H. (in press). From anecdotes to real-life examples: a mixed-methods survey on self-deception. *Cognition, Brain, Behaviour. An Interdisciplinary Journal*
Neal, T. M. S., Hight, M., Howatt, B. C., & Hamza, C. (2018). The cognitive and social psychological bases of bias in forensic mental health judgments. In M. K. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (pp. 151–175). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-75859-6_5
Newman, E. J., Swire-Thompson, B., & Ecker, U. K. H. (2022). Misinformation and the Sins of Memory: False-Belief Formation and Limits on Belief Revision. *Journal of applied research in memory and cognition*, *11*(4), 471–477. https://doi.org/10.1037/mac0000090
Ohlsson, I. M., & Ireland, J. L. (2011). Aggression and offence motivation in prisoners: exploring the components of motivation in an adult male sample. *Aggressive behavior, 37*(3), 278–288. https://doi.org/10.1002/ab.20386
Oliveira, C. M., & Levine, T. R. (2008). Lie acceptability: A construct and measure. *Communication Research Reports, 25*(4), 282-288. https://doi.org/10.1080/08824090802440170
Ost, J. (2017). Adults' retractions of childhood sexual abuse allegations: High-stakes and the (in)validation of recollection. *Memory*, *25*(7), 900–909. https://doi.org/https://doi.org/10.1080/09658211.2016.1187757
Ost, J., Foster, S., Costall, A., & Bull, R. (2005). False reports of childhood events in appropriate interviews. *Memory, 13*(7), 700–710. https://doi.org/10.1080/09658210444000340
Otgaar, H., & Baker, A. (2018). When lying changes memory for the truth. *Memory, 26*(1), 2–14. https://doi.org/10.1080/09658211.2017.1340286
Otgaar, H., Candel, I., & Merckelbach, H. (2008). Children's false memories: easier to elicit for a negative than for a neutral event. *Acta Psychologica*, *128*(2), 350–354. https://doi.org/10.1016/j.actpsy.2008.03.009

Otgaar, H., Candel, I., Merckelbach, H., & Wade, K. A. (2009). Abducted by a UFO: Prevalence information affects young children's false memories for an implausible event. *Applied Cognitive Psychology, 23*(1), 115-125.

Otgaar, H., Candel, I., Scoboria, A., & Merckelbach, H. (2010). Script knowledge enhances the development of children's false memories. *Acta psychologica*, *133*(1), 57–63. https://doi.org/10.1016/j.actpsy.2009.09.002

Otgaar, H., Curci, A., Mangiulli, I., Battista, F., Rizzotti, E., & Sartori, G. (2022a). A court ruled case on therapy-induced false memories. *Journal of Forensic Sciences*, *67*(5), 2122–2129. https://doi.org/10.1111/1556-4029.15073

Otgaar, H., Houben, S. & Howe, M. L. (2018). Methods of Studying False Memory. *Research methods in human memory.* Routledge.

Otgaar, H., Howe, M. L., Mangiulli, I., & Bücken, C. (2020). The impact of false denials on forgetting and false memory. *Cognition, 202*(104322), 1–9. https://doi.org/10.1016/j.cognition.2020.104322

Otgaar, H., Howe, M. L., Memon, A., & Wang, J. (2014). The development of differential mnemonic effects of false denials and forced confabulations. *Behavioral Sciences and the Law, 32*(6), 718–731. https://doi.org/10.1002/bsl.2148

Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus, E. F. (2019). The return of the repressed: the persistent and problematic claims of long-forgotten trauma. *Perspectives on psychological Science*, *14*(6), 1072–1095. https://doi.org/10.1177/1745691619862306

Otgaar, H., Howe, M. L., Peters, M., Sauerland, M., & Raymaekers, L. (2013). Developmental trends in different types of spontaneous false memories: implications for the legal field. *Behavioral sciences & the law*, *31*(5), 666–682. https://doi.org/10.1002/bsl.2076

Otgaar, H., Howe, M. L., Smeets, T., & Wang, J. (2016a). Denial-induced forgetting: False denials undermine memory, but external denials undermine belief. *Journal of Applied Research in Memory and Cognition, 5*(2), 168–175. https://doi.org/10.1016/j.jarmac.2016.04.002

Otgaar, H., Mangiulli, I., Battista, F., & Howe, M. L. (2023). External and internal influences yield similar memory effects: The role of deception and suggestion. *Frontiers in Psychology*, *14*, 1081528. https://doi.org/10.3389/fpsyg.2023.1081528

Otgaar, H., Mangiulli, I., Li, C., Jelicic, M., & Muris, P. (2025). The recovery and retraction of memories of abuse: A scoping review. *Frontiers in Psychology*, *16*, 1498258. https://doi.org/10.3389/fpsyg.2025.1498258

Otgaar, H., Moldoveanu, G., Melis, V., & Howe, M. L. (2022b). A new method to implant false autobiographical memories: Blind implantation. *Journal of Applied Research in Memory and Cognition, 11*(4), 580–586. https://doi.org/10.1037/mac0000028

Otgaar, H., Moldoveanu, G., Wang, J., & Howe, M. L. (2016b). Exploring the consequences of nonbelieved memories in the DRM paradigm. *Memory*, *25*(7), 922–933. https://doi.org/10.1080/09658211.2016.1272701

Otgaar, H., Riesthuis, P., Neal, T. M. S., Chin, J., Boskovic, I., & Rassin, E. (2023). If generalization is the grail, practical relevance is the nirvana: Considerations from the contribution of psychological science of memory to law. *Journal of Applied Research in Memory and Cognition, 12*(2), 176–179. https://doi.org/10.1037/mac0000116

Otgaar, H., Riesthuis, P., Ramaekers, J. G., Garry, M., & Kloft, L. (2022c). The importance of the smallest effect size of interest in expert witness testimony on alcohol and memory. *Frontiers in Psychology*, *13*, 980533. https://doi.org/10.3389/fpsyg.2022.980533

Otgaar, H., Romeo, T., Ramakers, N., & Howe, M. L. (2018). Forgetting having denied: the "amnesic" consequences of denial. *Memory and Cognition, 46*, 520–529. https://doi.org/10.3758/s13421-017-0781-5

Otgaar, H., Scoboria, A., Howe, M. L., Moldoveanu, G., & Smeets, T. (2016c). Challenging memories in children and adults using an imagination inflation procedure. *Psychology of Consciousness: Theory, Research, and Practice*, *3*, 270–283.

Otgaar, H., Smeets, T., & Peters, M. (2012). Children's implanted false memories and additional script knowledge. *Applied Cognitive Psychology, 26*(5), 709–715. https://doi.org/10.1002/acp.2849

Parker, A., & Morgan, H. (2020). Citizenship, marginalisation and youth offending: Acceptance, responsibility and resettlement. *Sociological Research Online, 25*(3), 507-523. https://doi.org/10.1177/1360780419888924

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46(3)*, 598–609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L., Curtis, S. R., & Jones, D. N. (2018). Aggression as a trait: The Dark Tetrad alternative. *Current Opinion in Psychology, 19*, 88-92. https://doi.org/10.1016/j.copsyc.2017.04.007

Paulhus, D. (1998). *Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding—7: User's manual*. Multi-Health System

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Lawrence Erlbaum Associates Publishers.

Paulhus, D. L. (2008). Self-deception. In R. F. Baumeister & K. D. Vohs(Eds.), *Encyclopedia of social psychology* (pp. 189–190). Sage Press

Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 227–246). Oxford University Press.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in Socially desirable responding. *Journal of Personality and Social Psychology, 60,* 307-317.

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556-563. https://doi.org/10.1016/S0092-6566(02)00505-6

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. Journal of Personality and Social Psychology, 84(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Perinelli, E., & Gremigni, P. (2016). Use of social desirability scales in clinical psychology: A systematic review. *Journal of Clinical Psychology, 72*(6), 534–551. https://doi.org/10.1002/jclp.22284

Peterson, J. B., DeYoung, C. G., Driver-Linn, E., Séguin, J. R., Higgins, D. M., Arseneault, L., & Tremblay, R. E. (2003). Self-deception and failure to modulate responses despite accruing evidence of error. *Journal of Research in Personality*, *37*(3), 205-223. https://doi.org/10.1016/S0092-6566(02)00569-X

Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development, 70*(4), 887–895. https://doi.org/10.1111/1467-8624.00064

Pezdek, K., & Freyd, J. J. (2009). The Fallacy of Generalising from Egg Salad in False-Belief Research. *Analyses of Social Issues and Public Policy*, *9*(1), 177-183. https://doi.org/10.1111/j.1530-2415.2009.01178.x

Pezdek, K., Finger, K., & Hodge, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science, 8*(6), 437 441. https://doi.org/10.1111/j.1467-9280.1997.tb00457.x

Polage, D. C. (2004). Fabrication Deflation? The Mixed Effects of Lying on Memory. *Applied Cognitive Psychology, 18*(4), 455–465. https://doi.org/10.1002/acp.995

Polage, D. C. (2012). Fabrication inflation increases as source monitoring ability decreases. *Acta Psychologica*, **139**(2), 335–342. https://doi.org/10.1016/j.actpsy.2011.12.007

and Behavior. *43*(10), 1293-1309. https://doi.org/10.1177/0093854815628027

Ren, M., Zhong, B., Fan, W., Dai, H., Yang, B., Zhang, W., Yin, Z., Liu, J., Li, J., & Zhan, Y. (2018). The Influence of Self-Control and Social Status on Self-Deception. *Frontiers in Psychology*, *9*, https://doi.org/10.3389/fpsyg.2018.01256

Riesthuis, P., Howe, M. L., & Otgaar, H. (2025). Meaningful approaches to assessing the size of effects in memory research: applications and recommendations for study design, interpretation, and analysis. *Memory*, 1–10. https://doi.org/10.1080/09658211.2025.2492496

Riesthuis, P., Mangiulli, I., Broers, N., & Otgaar, H. (2021). Expert opinions on the smallest effect size of interest in false memory research. *Applied Cognitive Psychology*, *36*(1), 203-215. https://doi.org/10.1002/acp.3911

Riesthuis, P., Otgaar, H., Battista, F., & Mangiulli, I. (2022). Public beliefs on the relationship between lying and memory. *Psychology, Crime & Law, 28*(6), 545-568, https://doi.org/10.1080/1068316X.2021.1929979

Riesthuis, P., Otgaar, H., Bogaard, G., & Mangiulli, I. (2023). Factors affecting the forced confabulation effect: a meta-analysis of laboratory studies. *Memory (Hove, England)*, *31*(5), 635–651. https://doi.org/10.1080/09658211.2023.2185931

Riesthuis, P., Otgaar, H., Hope, L., & Mangiulli, I. (2022). Registered report: The effects of incentivized lies on memory. *Applied Cognitive Psychology*, *36*(5), 1150-1161. https://doi.org/10.1002/acp.3986

Riesthuis, P., Otgaar, H., Mangiulli, I., & de Tauzia, Romane (2020). Adopting a fictitious autobiography: fabrication inflation or deflation?, *Memory*, *28*(6), 741-752, https://doi.org/10.1080/09658211.2020.1771371

Ritchie, T. D., Sedikides, C., & Skowronski, J. J. (2016). Does a person selectively recall the good or the bad from their personal past? It depends on the recall target and the person's favourability of self-views. *Memory, 25*(8), 934–944. https://doi.org/10.1080/09658211.2016.1233984

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology, 80*(2), 340–352. https://doi.org/10.1037/0022-3514.80.2.340

Robinson, M. D., & Ryff, C. D. (1999). The role of self-deception in perceptions of past, present, and future happiness. *Personality and Social Psychology Bulletin, 25*(5), 595–606. https://doi.org/10.1177/0146167299025005005

Robinson, M. D., Moeller, S. K., & Goetz, P. W. (2008). Are self-deceivers enhancing positive affect or denying negative affect? Toward an understanding of implicit affective processes. *Cognition and Emotion*, *23*(1), 152–180. https://doi.org/10.1080/02699930801961707

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roșan, A., Frick, P. J., Gottlieb, K. A., & Fasicaru, L. (2015). Callous-unemotional traits and anxiety in a sample of detained adolescents in Romania. *Journal of Evidence-Based Psychotherapies, 15*(1). 79-95. https://repository.lsu.edu/psychology_pubs/390/

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology, 46*(2), 237–248. https://doi.org/10.1037/0022-3514.46.2.237

Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology, 47*(1), 213–215. https://doi.org/10.1037/0022-006X.47.1.213

Sackeim, H.A., Gur, R.C. (1978). Self-Deception, Self-Confrontation, and Consciousness. In: Schwartz, G.E., Shapiro, D. (eds*) Consciousness and Self-Regulation*. Springer. https://doi.org/10.1007/978-1-4684-2571-0_4

Sagana, A., Sauerland, M., & Merckelbach, H. (2014). Memory impairment is not sufficient for choice blindness to occur. *Frontiers in Psychology*, *5*, 87-140.

Schacter, D. L., Greene, C. M., & Murphy, G. (2023). Bias and constructive processes in a self-memory system. Memory, 32(6), 656–665. https://doi.org/10.1080/09658211.2023.2232568

Schindler, S., Wagner, L. K., Reinhard, A., Ruhara, N., Pfattheicher, S., & Nitschke, J. (2020). Are criminals better lie detectors? Investigating offenders' abilities in the context of deception detection. *Applied Cognitive Psychology, 35(1),* 203-214. https://doi.org/10.1002/acp.3755

Schnapp P., Eggert S. & Suhr R. (2017). Comparing Continuous and Dichotomous Scoring of Social Desirability Scales: Effects of Different Scoring Methods on the Reliability and Validity of the Winkler-Kroh-Spiess BIDR Short Scale. *Survey Methods: Insights from the Field*. Retrieved from https://surveyinsights.org/?p=8708

Schwardmann, P., & van der Weele, J. (2019). Deception and self-deception. *Nature human behaviour*, *3*(10), 1055–1061. https://doi.org/10.1038/s41562-019-0666-7

Scoboria, A., Jackson, D. L., Talarico, J., Hanczakowski, M., Wysman, L., & Mazzoni, G. (2014). The role of belief in occurrence within autobiographical memory. *Journal of Experimental Psychology. General*, *143*(3), 1242–1258. https://doi.org/10.1037/a0034110

Scoboria, A., Mazzoni, G., Kirsch, I., & Relyea, M. (2004). Plausibility and belief in autobiographical memory. *Applied Cognitive Psychology, 18*(7), 791–807. https://doi.org/10.1002/acp.1062 Strange, D., Sutherland, R., & Garry, M. (2006). Event plausibility does not determine children's false memories. *Memory, 14*(8), 937 -951. https://doi.org/10.1080/09658210600896105

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2016). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, *25*(2), 146–163. https://doi.org/10.1080/09658211.2016.1260747

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, *25*(2), 146–163. https://doi.org/10.1080/09658211.2016.1260747

Sedikides, C., & Alicke, M. D. (2012). Self-enhancement and self-protection motives. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 303–322). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0017

Sedikides, C., & Green, J. D. (2000). On the self-protective nature of inconsistency-negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of Personality and Social Psychology, 79*(6), 906–922. https://doi.org/10.1037/0022-3514.79.6.906

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3*(2), 102–116. https://doi.org/10.1111/j.1745-6916.2008.00068.x

Sedikides, C., & Skowronski, J. J. (1996). The Symbolic Self in Evolutionary Context. *Personality and Social Psychology Review*. https://doi.org/10.1207/s15327957pspr0101_6

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*(1), 60–79. https://doi.org/10.1037/0022-3514.84.1.60
Sedikides, C., Gaertner, L., & Vevea, J. L. (2005). Pancultural self-enhancement reloaded: A meta-analytic reply to Heine (2005). *Journal of Personality and Social Psychology, 89*(4), 539–551. https://doi.org/10.1037/0022-3514.89.4.539
Seidl, H., Nilsson, T., Hofvander, B., Billstedt, E., & Wallinius, M. (2020). Personality and Cognitive Functions in Violent Offenders - Implications of Character Maturity?. *Frontiers in Psychology, 11*, 58. https://doi.org/10.3389/fpsyg.2020.00058
Seitz, R. J., & Angel, H. (2020). Belief formation – A driving force for brain evolution. *Brain and Cognition*, *140*, 105548. https://doi.org/10.1016/j.bandc.2020.105548
Sengupta, P., Das, S., Bhattacharyya, J., Ray, D., & Palit, R. G. (2020). Exploring the underlying psychological correlates of self-deception. *Indian Journal of Health and Wellbeing, 11*(1–3), 18–22.
Serota, K. B., & Levine, T. R. (2015). A few prolific liars: Variation in the prevalence of lying. *Journal of Language and Social Psychology, 34*(2), 138–157. https://doi.org/10.1177/0261927X14528804
Serota, K. B., Levine, T. R., & Docan-Morgan, T. (2021). Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both? *Communication Monographs, 89*(3), 307–331. https://doi.org/10.1080/03637751.2021.1985153
Serota, K. B., Levine, T. R., Zvi, L., & Markowitz, D. M. (2024). The ubiquity of long-tail lie distributions: Seven studies from five continents. *Journal of Communication, 74*(1), 1-11. https://doi.org/10.1093/joc/jqad040
Shaw, J., & Vredeveldt, A. (2019). The Recovered Memory Debate Continues in Europe: Evidence From the United Kingdom, the Netherlands, France, and Germany. *Clinical Psychological Science, 7*(1), 27-28. https://doi.org/10.1177/2167702618803649
Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, *2*(2), 895-908. https://doi.org/10.1111/j.1751-9004.2008.00078.x
Sherman, D. K., & Cohen, G. L. (2005). The psychology of self-defence: self-affirmation theory. *Advances in Experimental Social Psychology, 38,* 183-242. https://doi.org/10.1016/S0065-2601(06)38004-5
Sharman, S. J., & Scoboria, A. (2009). Imagination equally influences false memories of high and low plausibility events. *Applied Cognitive Psychology*, *23*(6), 813-827. https://doi.org/10.1002/acp.1515
Shahvaroughi, A., Dyevre, A., & Otgaar, H. (2025). Examining the effects of mood and emotional valence on the creation of false autobiographical memories. *Memory & cognition*, 10.3758/s13421-025-01697-x. Advance online publication. https://doi.org/10.3758/s13421-025-01697-x
Sheridan, Z., Boman, P., Mergler, A., & Furlong, M. J. (2015). Examining well-being, anxiety, and self-deception in university students. *Cogent Psychology*, *2*(1). https://doi.org/10.1080/23311908.2014.993850
Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330-349. https://doi.org/10.1177/0146167211398138
Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, *115*(2), 268- 281. https://doi.org/10.1016/j.cognition.2009.12.017
Smeets, T., Merckelbach, H., Horselenberg, R., & Jelicic, M. (2005). Trying to recollect past events: Confidence, beliefs, and memories. *Clinical Psychology Review*, *25*(7), 917-934. https://doi.org/10.1016/j.cpr.2005.03.005
Smith, M. E. (2007). Self-Deception Among Men Who Are Mandated to Attend a Batterer Intervention Program. *Perspectives in Psychiatric Care, 43*(4), 193-203. https://doi.org/10.1111/j.1744-6163.2007.00134.x
Smith, M. K., Trivers, R., & Von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, *63*, 93-101. https://doi.org/10.1016/j.joep.2017.02.012
Stake, J. E. (1994). Development and validation of the six-factor self-concept scale for adults. *Educational and Psychological Measurement, 54*(1), 56-72. https://doi.org/10.1177/0013164494054001006
Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing Continuous and Dichotomous Scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*(2), 370–389. https://doi.org/10.1207/S15327752JPA7802_10
Stanley, M., Yang, B., & De Brigard, F. (2018). No evidence for unethical amnesia for imagined actions: a failed replication and extension. https://doi.org/10.31234/osf.io/kn8ce
Strange, D., Sutherland, R., & Garry, M. (2006). Event plausibility does not determine children's false memories. *Memory, 14*(8), 937 -951. https://doi.org/10.1080/09658210600896105
Steinert T. (2002). Prediction of inpatient violence. *Acta Psychiatrica Scandinavica. Supplementum, 106,* 133-141. https://doi.org/10.1034/j.1600-0447.106.s412.29.x
Swann, W. B., Jr., & Buhrmester, M. D. (2012). Self-verification: The search for coherence. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (2nd ed., pp. 405–424). The Guilford Press.

Tan, L., & Grace, R. C. (2008). Social Desirability and Sexual Offenders: A Review. *Sexual Abuse, 20*(1), 61-87. https://doi.org/10.1177/1079063208314820

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Portrait of the self-enhancer: well-adjusted and well-liked or maladjusted and friendless?. *Journal of personality and social psychology*, *84*(1), 165–176.

Tice, D. M., & Masicampo, E. J. (2008). Approach and avoidance motivations in the self-concept and self-esteem. In A. J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 505–519). Psychology Press.

Tomaka, J., Blascovich, J., & Kelsey, R. M. (1992). Effects of Self-Deception, Social Desirability, and Repressive Coping on Psychophysiological Reactivity to Stress. *Personality and Social Psychology Bulletin*, *18*(5), 616-624. https://doi.org/10.1177/0146167292185012

Trivers, R. (2000). The elements of a scientific theory of self-deception. In D. LeCroy & P. Moller (Eds.), *Evolutionary perspectives on human reproductive behavior* (pp. 114–131). New York Academy of Sciences.

Uner, O., & Roediger, H. L. (2022). Do recall and recognition lead to different retrieval experiences? *American Journal of Psychology, 135* (1), 33–43. https://doi.org/10.5406/amerjpsyc.135.1.0033

Vigil-Colet, A., Ruiz-Pamies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema, 24*(2), 310-315. https://pubmed.ncbi.nlm.nih.gov/22420362/

Visu-Petra, G., Jurje, O., Opre, A., Miclea, M., & Visu-Petra, L. (2014). Which came first, the truth or the lie? Effects of order and habituation in the reaction time-based concealed information test. *Cognition, Brain, Behavior: An Interdisciplinary Journal*, **18**(4), 283–297.

Vrij, A., Mann, S., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior, 32*(3), 253-265. https://doi.org/10.1007/s10979-007-9103-y

Vrij, A. (2008). Detecting lies and deceit: Pitfalls and opportunities. John Wiley & Sons. https://psycnet.apa.org/record/2008-01237-000

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *The Behavioral and Brain Sciences*, *34*(1), 1–56.

von Hippel, W., Lakin, J. L., & Shakarchi, R. J. (2005). Individual Differences in Motivated Social Cognition: The Case of Self-Serving Information Processing. *Personality and Social Psychology Bulletin*, *31*(10), 1347-1357. https://doi.org/10.1177/0146167205274899

Wade, K. A., M. Garry, & Pezdek K..(2018). Deconstructing rich false memories of committing crime: commentary on Shaw and Porter (2015). *Psychological Science 29,* (3), 471–476. https://doi.org/10.1177/0956797617703667

Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: using false photographs to create false childhood memories. *Psychonomic bulletin & review*, *9*(3), 597–603. https://doi.org/10.3758/bf03196318

Wade, K. A., Riesthuis, P., Bücken, C., Otgaar, H., & Loftus, E. F. (2025). Still lost in the mall—false memories happen and that's what matters. *Applied Cognitive Psychology, 39*(1), e70028. https://doi.org/10.1002/acp.70028

Wang, J., Otgaar, H., Bisback, A., Smeets, T., & Howe, M. L. (2019a). The consequences of implicit and explicit beliefs on food preferences. *Psychology of Consciousness: Theory, Research, and Practice, 6*(4), 371–385. https://doi.org/10.1037/cns0000203

Wang, J., Otgaar, H., Smeets, T., Howe, M. L., & Zhou, C. (2019b). Manipulating memory associations changes decision-making preferences in a preconditioning task. *Consciousness and Cognition*, *69*, 103-112. https://doi.org/10.1016/j.concog.2019.01.016

Wang, J., Otgaar, H., Santtila, P., Shen, X., & Zhou, C. (2021). How culture shapes constructive false memory. *Journal of Applied Research in Memory and Cognition, 10*(1), 24–32. https://doi.org/10.1037/h0101792

Wang, J., Wang, B., Otgaar, H., Patihis, L., & Sauerland, M. (2024). Self-relevance enhances susceptibility to false memory. *Behavioral sciences & the law*, *42*(2), 79–95. https://doi.org/10.1002/bsl.2644

Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology*, *88*(4), 369–380. https://doi.org/10.1037//0021-843x.88.4.369

Westland, S., & Shinebourne, P. (2009). Self-deception and the therapist: An interpretative phenomenological analysis of the experiences and understandings of therapists working with clients they describe as self-deceptive. *Psychology and Psychotherapy: Theory, Research and Practice, 82*(4), 385- 401. https://doi.org/10.1348/147608309X450508

Wojcik, S. P., & Ditto, P. H. (2014). Motivated happiness: Self-enhancement inflates self-reported subjective well-being. *Social Psychological and Personality Science, 5*(7), 825–834. https://doi.org/10.1177/1948550614534699

Wright, D. B., Memon, A., Skagerberg, E. M., & Gabbert, F. (2009). When Eyewitnesses Talk. *Current Directions in Psychological Science*, *18*(3), 174-178. https://doi.org/10.1111/j.1467-8721.2009.01631.x

Wright, G. R.T., Berry, C. J., Catmur, C., & Bird, G. (2015). Good Liars Are Neither 'Dark' Nor Self-Deceptive. *PLOS ONE*, *10*(6), e0127315. https://doi.org/10.1371/journal.pone.0127315

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *The Academy of Management Review, 12*(2), 250–264. https://doi.org/10.2307/258533

Zhang, Y., Otgaar, H., and Wang, J. (2022b). Memory distrust is related to memory errors, self-25 esteem, and personality. *Appl. Cogn. Psychol. 36*, 283–292. 10.1002/acp.3917

Zhang, Y., Otgaar, H., Nash, R. A., & Rosar, L. (2024a). Time and memory distrust shape the dynamics of recollection and belief-in-occurrence. *Memory*, *32*(4), 484–501. https://doi.org/10.1080/09658211.2024.2336166

Zhang, Y., Otgaar, H., Riesthuis, P., Wang, J., & Jelicic, M. (2024b). The rise and fall of autobiographical beliefs: the effect of external feedback on memory in the context of the prisoner's dilemma. *Psychology of Consciousness: Theory, Research, and Practice*. Advance online publication. https://doi.org/10.1037/cns0000391

Zhang, Y., Pan, Z., Li, K., & Guo, Y. (2018). Self-serving bias in memories: Selectively forgetting the connection between negative information and the self.*Experimental Psychology, 65*(4), 236–244. https://doi.org/10.1027/1618-3169/a000409