STUDIA UBB PSYCHOL.-PAED., LXX, 1, 2025, p. 121-151
(Recommended Citation)
DOI:10.24193/subbpsyped.2025.1.06

Self-deception Beyond Speculation: A Narrative Review of the Empirical Research on Motivated False Beliefs

Mara Georgiana MOLDOVEANU[1,2], **Laura VISU-PETRA**[1,3*]

ABSTRACT. This narrative review explores the concept of self-deception, departing from its theoretical foundations in philosophy and psychology, and focusing on the pioneering empirical methods used to study it. We first outline key philosophical debates surrounding intentionality and paradoxes surrounding the concept of self-deception and then discuss influential psychological theories. The review covers major paradigms for measuring self-deception, such as the retrospective and forward-looking paradigms for situational self-deception and approaches focused on self-deception as a response bias, trait, or disposition. Our primary aim is to present the outcomes of the limited body of empirical research investigating motivational factors in self-deception, rather than theoretical speculations. We examine studies on both internal motivations (e.g., maintaining self-concept) and external motivations (e.g., deceiving others), highlighting how different motivational contexts influence the likelihood and extent of self-deception. Finally, we examine potential limitations, explore future research directions, and consider the broader implications of focusing on this particular aspect of self-deception, the motivational reasoning.

Keywords: self-deception, motivation, motivated false beliefs, external motivation, internal motivation

[1] Research in Individual Differences and Legal Psychology (RIDDLE) Lab, Department of Psychology, Babeş-Bolyai University, Cluj-Napoca, Romania.
[2] Forensic Psychology Section, Faculty of Psychology and Neuroscience, Maastricht University.
[3] Department of Social and Human Research, Romanian Academy, 400015 Cluj-Napoca, Romania.
[*] Corresponding authors: laurapetra@psychology.ro; maramoldoveanu@psychology.ro

"Do you see over yonder, friend Sancho, thirty or forty hulking giants? I intend to do battle with them and slay them."
Miguel de Cervantes Saavedra, *Don Quixote* (Pass, 2024)

INTRODUCTION

Self-deception is a complex phenomenon studied across multiple disciplines, including literature, philosophy, psychology, evolutionary biology, and behavioural economics (Mele, 2001; von Hippel & Trivers, 2011; Chance et al., 2011). Psychologists' interest in this phenomenon followed anecdotal evidence of psychoanalytical therapy cases (e.g., rationalisation, repression, projections and denial, Freud, 1960; an idea still present in contemporary research, see Westland & Shinebourne, 2009). According to psychoanalytic views, self-deception is an intrapersonal, adaptive defence mechanism in which individuals engage to avoid thinking of painful truths (Johnson, 1995). Over the years, some authors proposed the overlap of two concepts (e.g., repression and self-deception or denial and self-deception, Baumeister & Cairns, 1992), while other authors attempted to differentiate self-deception from other allegedly defensive mechanisms (e.g., repression, Gur & Sackeim, 1979; see also Asley & Holtgraves, 2003; Tomaka et al., 1992).

During the past decades, pioneering empirical psychological articles have been published, shedding some light on possible mechanisms and functions (Chance et al., 2011; Chance & Norton, 2015; Liu et al., 2025; Mei et al., 2023; Ren et al., 2018; Mijović-Prelec & Prelec, 2010; Smith et al., 2017). However, significant differences among scholars regarding its precise definition and underlying mechanisms remain (Bachkirova, 2016; Balcetis, 2008; Khalil, 2017; Mele, 1997; von Hippel & Trivers, 2011; Paulhus & Holden, 2010; Fan et al., 2022).

Furthermore, psychoanalytic constructs such as repression have been at the centre of a heated scientific debate due to their implications for memory recovery in abuse cases. For instance, the beliefs of therapists in repression have been linked with false memories, which could have severe consequences in the legal field (see Battista et al., 2023; Otgaar et al., 2019). However, studies on self-deception are yet to address fundamental aspects of the phenomenon, such as the systematic investigation of mechanisms that could lead to the successful formation of self-deception (Mele, 2000; Sackeim & Gur, 1979; von Hippel & Trivers, 2011).

Additionally, a substantial part of scientific literature on self-deception consists of theoretical and essentially speculative postulates (e.g., Balcetis, 2008; Mele, 2001; Lauria et al., 2016; von Hippel & Trivers, 2011). These

proposals focus on different ways of understanding the phenomenon of 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).

The paucity of concrete empirical evidence of self-deception highlights a significant gap in understanding this concept. Furthermore, 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) of self-deception contributes to its limited conceptual validity despite its importance in everyday life, and especially in legal contexts where false beliefs could lead to judicial errors.

From Philosophical to Psychological Perspectives on Self-deception

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). Initially, more influential philosophical theories of self-deception were modelled on intrapersonal deception. These attempts at defining and explaining self-deception led to two paradoxes (Mele, 1997, 2001). The static paradox of how the self-deceived individual would have to hold simultaneously two contradictory, incompatible beliefs (e.g. thinking that "loyalty in romantic relationships is very important for me", by a person who cheated on their partner repeatedly). The dynamic paradox is more related to the process of self-deception, rather than a product, and it refers to the process through which a person intentionally acquires a belief while remaining unaware of that belief. To address these paradoxes, 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 static and dynamic paradoxes are addressed using the psychoanalytic perspective, which divides the self 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, 2010). 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), the static and dynamic paradoxes are denied. Instead, 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 and there is no separate "deceiver" and "deceived" in our mind, suggesting that at no moment must the individual believe in two contradictory beliefs at the same time. The minimal conditions that deflationary philosophers agreed upon were that "self-deception involves an individual who acquires and maintains some beliefs despite contrary evidence and who may display behaviours suggesting some awareness of the truth, as a consequence of some motivation" (Mele, 2001; Sackeim & Gur, 1979).

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; 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 in order to better deceive others (for example, being a byproduct of other-deception; Schwardmann & van der Weele, 2019; von Hippel & Trivers, 2011). Some authors propose that the two perspectives do not have to be mutually exclusive or situated at opposite ends of a spectrum; rather, the proposed intrapersonal and interpersonal motivations and mechanisms could be interdependent (for example, financial gains and increased confidence; Mijović-Prelec & Prelec, 2010).

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 construct, 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). The more negative perspective of self-deception as denial, repression, and self-protection has been more at the centre of the psychological literature at the end of the past century (Baumeister, 1996; Sackeim & Gur, 1978). During the last two decades, the focus has been more on the positive, self-enhancing dimension of self-deception (Chance et al., 2011; 2015; von Hippel & Trivers, 2011; Liu et al., 2019; Mei et al., 2023).

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). While focusing on the process of self-deception is important, some authors point out that for self-deception to be achieved, subjects must "lie to themselves successfully", meaning they have to convince themselves that something false is true (Paulhus, 2007; Sloman et al., 2010).

Beyond theoretical definitions, the psychological literature focused for the first time on operationalising self-deception for conducting empirical studies (e.g., Sackeim & Gur, 1979). While many paradigms have been developed or proposed, we will summarise below some of the most pivotal perspectives developed in psychological research during the past decades.

Operationalization and Measurement of Self-deception in Psychological Research

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 & Sackeim (1979) conducted one of the first *experiments* in the field to elicit self-deception and prove its existence. This study investigated the coexistence of two incompatible beliefs by having participants identify and categorise a speaker's voice (their own voice vs. others' voices; all previously recorded). The voice recognition task was accompanied by physiological indications (galvanic skin response). When the participants' verbal assessment, 'this is not my own voice,' conflicted with their physiological assessment showing reactions that could be interpreted as recognition of one's voice, this was taken as evidence of self-deception. Furthermore, participants were given either "success" or "fail" feedback before the voice identification task. The researchers predicted that the "fail" group would commit more misidentifications of self, while the "success" group would commit more misidentifications of others. The results, which were in line with the predictions made, were interpreted by the authors as confirming the hypothesis that self-deception was influenced by the tendency to minimise the discomfort (i.e., cognitive discrepancy) felt by the participants, making the case for motivated false beliefs.

Another early pioneering attempt at eliciting self-deception was made by Quattrone & Tversky (1984). They conducted an experiment where subjects were asked to keep their hands submerged in a cold water container until they

could no longer tolerate the pain. Afterward, during a debriefing, participants were informed that a certain inborn heart condition could be diagnosed by the effect of exercise on cold tolerance and told that "the consequences of this condition included a shorter lifespan and reduced quality of life". Some subjects were told that having a bad heart would increase cold tolerance, while others were informed of the opposite (including charts designed with the same info). After the debriefing, the subjects were told to exercise for a minute and then to repeat the cold water tolerance test. The majority of the participants showed changes correlated with the "good news" and "cheating on their own diagnostic". In a similar study, Kunda (1987) asked participants (women) to evaluate the credibility of a (fake) study linking coffee consumption and breast cancer, and results showed that self-reported heavier drinkers of coffee were more critical and less persuaded by the evidence provided. This suggested that participants kept their attitudes to match their previous behaviours despite receiving contradictory evidence. Thus, it was concluded that the favoured beliefs receive preferential treatment in information processing. A more recent study (Fernbach et al., 2014) using this paradigm, showed that when asked about the effort task, participants misrepresented the effort they put into the task in a self-serving way, minimising it when it was offering them a good diagnostic (e.g., skin quality), which is different than the effort they actually put in.

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). However, these attempts to explain self-deception were also the subject of criticism. One main critique was that, for example, physiological data or observed behaviour should not be assumed to directly equate to the presence of a belief (Mijović-Prelec & Prelec, 2010). Moreover, 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). This critique might partially explain why the initial empirical approaches to eliciting self-deception were not significantly pursued later in the literature.

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 being offered 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 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).

To sum up, most experimental paradigms looking at situational self-deception focus on the inconsistency between past behaviour in a real event (e.g., a test; seen as contradictory evidence) and the individual's evaluation or estimation of past or future experiences (e.g., tested as reaction, behaviour or self-assessment, thus the indication or the false belief).

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).

Paulhus and Reid (1991) further distinguished between self-deceptive denial and self-enhancement, with self-deceptive denial characterised by the denial of negative information and self-enhancement by the tendency to inflate positive characteristics related to oneself. However, the Self-deceptive Denial Scale (SDD) of the BIDR measure was later dropped as psychometric properties showed that the items contained were more consistent with the other construct of the measure, construct which is directed more towards other deception, namely impression management (a conscious attempt at showing a socially desirable image towards others, Paulhus, 2002). The Self-Enhancement Scale contains items worded positively and negatively (thus reflecting also the denial aspect), summarised in a total self-deceptive enhancement score (Paulhus, 2002).

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; Hilderbrand 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 Hilderbrand, et al., 2018; Vigil-Colet et al., 2012).

The role of self-deception (and social desirability) in measuring other constructs is often difficult to assess due to underreported findings about how this measure was used. For example, studies reporting only that their final results were controlled for response bias, but not the actual steps or studies frequently reporting a single total score for social desirability. However, there is evidence suggesting that the two components of social desirability, self-deceptive enhancement, and impression management, may have distinct associations with the outcomes and certain factors differentially influencing them (e.g., in the relation between dynamic risk factors, sample size, or setting (e.g., incarcerated) selectively influencing self-enhancement, but not impression management (Hildebrand et al., 2018; Tan & Grace, 2008). Furthermore, recent research has shown that impression management is more susceptible to context than self-deceptive enhancement (e.g. when assessing aggression in a prison setting, Mills & Kroner, 2005). The latter is proposed to be more related to general universal aspects of one's self-image. However, when talking about culture, it is important to emphasise that self-deceptive enhancement was found to be sensitive to culture. For example, East Asian countries show less self-enhancement than WEIRD (Western Educated Industrialized Rich Democratic) countries (Hampton & Varnum, 2018).

Using these 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. 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). However, studies are scattered and often treat motivations separately, making it difficult to have an overview. Our focus for the next section is to look at some examples of empirical studies that investigate different motivations and to portray a more comprehensive picture of these findings.

Self-deception as Motivated False Beliefs: Empirical Evidence

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, 2010).

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 with regard to self-deception. Preceding narrative reviews presented an overview of some of these motivations and empirical findings, however, the focus was on interpersonal and external motivations (for example, Chance & Norton, 2015; von Hippel & Trivers, 2011).

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).

The following sections of this review will examine empirical research on motivations for self-deception, beginning with external ones and then exploring internal motivations (see Table 1 for an overview of the proposed motivations in psychological research).

Table 1. *External and Internal Motivations for Self-Deception in Psychological Research*

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

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; von Hippel & Trivers, 2010; Smith et al., 2017).

Financial Incentives

Self-deception has been successfully elicited in experimental studies using financial incentives. Many paradigms studying deception and self-deception relied on monetary incentives to increase the deceptive behaviours in the laboratory (Chance et al., 2011, 2015; Lu et al., 2014; Mijović-Prelec & Prelec, 2010; Peterson et al., 2003). However, most studies do not primarily focus on financial incentives, and only a few studies have directly examined the role of monetary rewards and self-deception.

For example, the anticipation of a financial gain led participants (N = 85, students) in Mijović-Prelec and Prelec (2010) to engage in 20% more instances of self-deception in a categorisation task (abstract categorisation of drawings). In this study, self-deception was operationalised as an adjustment of a categorisation decision based on a previously anticipated categorisation. More specifically, participants rated a Korean written sign (character) as appearing to be more

male-like and were asked to anticipate what character would be next shown. When they predicted that a female-like-looking character would follow, and were offered the opportunity to receive an additional incentive, given only when the anticipation was "confirmed", they were more likely to rate the next character as actually female-like, even if before (in the initial categorisation task) they qualified the same character as male-like. Thus, participants adjusted their categorisation decisions (self-deception) to align with their anticipated outcomes to gain more financial incentives.

In another study, using the forward-thinking paradigm, Chance et al. (2011) explored whether using financial incentives for accurate future estimations would lead to a decrease in self-deception, thus less inflated performance estimations. They found that participants continued to show self-deceptive behaviours even when this led to not gaining money. Similarly, in a study by Peterson et al. (2003), 13-year-old boys and young adults participated in gambling-type card tasks in which higher performance led to more monetary gains. Looking at trait self-deceptive enhancement, participants were split into low and high self-deceivers. High self-deceivers played more cards and won significantly less money, even though they received evidence of error (they were informed that 19 of the last 20 cards were losing).

Taken together, these results suggest that when offered monetary benefits which favour deception (and self-deception), self-deception increases. However, self-deception does not decrease when financial incentives are given for the truth. While not focused on self-deception directly, in another study that looked at overclaiming participants and financial incentives, Mazaar et al. (2008) manipulated monetary amounts and types of financial gains (e.g., money vs. tokens). This study showed that individuals were dishonest to a limited extent and that the amount of dishonesty was not as much influenced by more financial incentives, thus raising the question of the limitations when engaging in self-deception for monetary gains.

Deceiving or Persuading Others

The postulate that self-deception has evolved to better deceive others has been proposed by Trivers (Trivers, 2000; von Hippel & Trivers, 2011). A few empirical studies tested this theory, showing some initial validation. For example, Smith et al. (2017) tested the hypothesis that self-deception evolved in service of persuasion and conducted an experiment in which participants were asked to write a persuasive text about an individual who behaved in a manner that was initially consistent or inconsistent with their persuasive goals (to convince others that a person was likeable or dislikeable). Participants were

shown brief videos depicting a person engaging in positive (e.g. helping someone in distress), neutral (e.g. making lunch), or negative (e.g. stealing money) behaviours. Furthermore, they were told that if their text is rated as persuasive by others, they will earn more money, and they also gave their private opinion of the person depicted in the videos. The results supported their hypothesis, thus, the participants biased their information-gathering strategies (gathering less information and the information that supports their view) and convinced themselves of its veracity, being more efficient in convincing others when they themselves believed it.

In another study, Lamba & Nityananda (2014) looked at self-deception and the comparison between participants' self-evaluations, others' evaluations, and actual behaviours. They tested if individuals' false beliefs about their abilities can influence how they are perceived by others. In their study, participants were college students (N = 73) for whom their performance and their self-evaluations versus the evaluation they received from their peers regarding their abilities were measured at the beginning and end of a semester. In this study, it was shown that overconfident individuals were more convincing and judged as better rated than underconfident individuals, who were judged less able than they actually were. However, this effect decreased in time and with extended interactions, showing that as individuals' levels of self-deception changed, their peers' judgments of them also changed. Additionally, participants' self-deceptive predictions about themselves were not correlated with their actual performance. This decrease in time and with extended interactions was also found by Dufner et al. (2019) in a meta-analysis, which looked at dispositional self-deception and interpersonal adjustment across multiple studies.

Another study (Wright et al., 2015) which looked at trait self-deception (using the self-deceptive enhancement scale described in the previous section) showed that high self-deceptive individuals (using a median split) were less credible and less confident when lying during a deception detection paradigm (in which participants have to lie and tell the truth regarding some of their previously recorded social attitudes in front of a group and rate how convinced they were of others lies).

Using another method and perspective (focusing on memory), Lu & Chang (2014) tested what happens with the recognition of studied items when participants (Chinese college students) had to deceive an equal or a higher-status individual (e.g., a teacher). They tested their memory for the previously studied items after the interaction with the deceived target, when the participant was alone. The results showed that when interacting with a high-status, but not with an equal-status individual, participants remembered fewer previously studied items than on a second memory test alone. These findings

suggest that during a situation in which one is more concerned about the consequences of being caught, one might temporarily self-deceive, suggesting that the theory of self-deception to deceive others might be supported.

Social Recognition and Validation

In the case of social recognition, a few studies also tried to examine if receiving social recognition or validation might lead to an increase in self-deception.

One of the first studies investigating the effect of social context on self-presentation strategies and reactions to feedback by participants was focused on "repressors" (operationalized as participants high in social desirability and low on anxiety), a construct which was seen as an equivalent to self-deceivers (Baumeister & Cains, 1992). Findings of this study showed that when being exposed to others' opinions, repressors were more concerned about their image, and recalled more of the negative feedback than when they received feedback privately. Participants were more susceptible to social validation, which led to a decrease in their ability to ignore negative feedback, attributed to engaging in self-deceiving strategies.

In another study, Chance et al. (2011) offered participants the possibility of receiving a certificate for their performance on a test and found that the participants who had this opportunity had more inflated self-estimations for a hypothetical future performance, thus higher self-deception. Similarly, Yang et al. (2024) gave participants social comparison feedback regarding their performances (e.g., you performed better or worse than your colleagues). Participants told that they performed worse than others showed higher self-deception (in estimating their future performance) compared with the participants being told they performed better (thus given self-affirming feedback). These results suggest that social validation and recognition might be a motive for situational self-deception.

Looking at trait or dispositional self-enhancement, in a meta-analytic review, Dufner et al. (2019) showed that self-enhancement was associated with a more positive assessment by others during a first meeting. However, this effect faded in the long term. Similarly, Lamba and Nityananda (2014) showed that individuals who rated themselves higher in their performance (e.g., university subjects) were also rated higher by others. However, this effect decreased over time, with extended interactions, showing that as individuals' levels of self-deception change, their peers' judgments of them also change. Thus, while social recognition and validation might be a motivation or associated with self-deception, this effect or benefit might be present in the short term and fade in the long run.

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; Maazar 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, 2010).

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.

Increased Confidence in Oneself

The tendency to be overconfident in our everyday abilities, skills, and personal traits by attributing success to internal factors and failure to external factors has long been documented (Miller & Ross, 1975) under the term "self-serving bias". It has been widely accepted as a fundamental need to maintain positive beliefs about oneself (Sedikides & Alicke, 2012), but see also self-verification theory, which posits that people search for information coherent with their self-view (Swann & Buhrmester, 2012).

Von Hippel & Trivers (2011) postulate that self-deception is associated with a higher confidence in one's abilities, which supports successful deception. However, only a few studies have investigated self-deception, especially situational, and measured confidence separately. For example, Mijović-Prelec & Prelec (2010) have shown that self-deception leads to increased confidence ratings (their confidence in their choice in a categorisation task), however, this increase in confidence was present only in the case of moderate self-deceivers. When the self-deception was higher in individuals, confidence decreased following contradictory evidence.

In another study, Schwardmann & van der Weele (2018) investigated what would happen if participants were offered the opportunity to receive gains from deceiving others; would this lead to overconfidence in one's performance and make the more overconfident participants more convincing to others? Their results showed these effects, thus supporting the theory that overconfidence is

related to self-deception. However, another study where self-deception was measured as a trait (Wright et al., 2015) showed that self-deceptive individuals were less credible and less confident when lying. These differences could be due to different measures and procedures used to measure self-deception (situational vs trait).

Psychological Adjustment, Subjective Well-Being, and Affective States

When looking at self-deception or self-deceptive enhancement and *subjective well-being* a positive association with personal adjustment or positive characteristics is usually reported, plus a negative association with unfavourable characteristics or states (anxiety, depression, etc.) (Baumeister, 1993; Duftenr et al., 2019; Robinson and Ryff, 1999; Taylor et al., 2003). As early as 1989, Taylor (1989) noted that individuals who made overly positive self-evaluations had higher self-esteem, reported greater happiness, showed a heightened ability to care for others, and experienced an enhanced capacity for creative and productive work. Thus, self-deception was proposed to serve as a buffer against negative emotional states, enhancing subjective well-being (Taylor et al., 2003).

To date, research findings mostly on dispositional self-deception indicate that individuals exhibiting higher levels of self-deception tend to report greater overall positive psychological traits. For example, a study by Erez et al. (1995) showed that trait self-deception mediated the relationship between affective dispositions, locus of control, and subjective well-being, both when reported by self as when reported by others (N = 219 self-report and N = 211 self and significant others' reports). In a more recent meta-analysis by Dufner and colleagues (2019), which examined the association between self-enhancement and personal adjustment, self-enhancement was positively and robustly related to personal adjustment across sex, age, cohort, and culture. Another study showed that individuals tend to view their future selves more positively, with self-deception contributing uniquely to predicting future happiness, but not past happiness evaluations (Robinson & Ryff, 1999).

However, while self-deceptive enhancement accounts for a percentage of the variance in psychological subjective well-being, when seen as a trait or disposition, it could also be these associations are due to self-enhancement being a confounded variable meaning a positive response tendency that leads to inflated scores on well-being measures (Wojcik & Ditto, 2014). To our knowledge, the relationship between subjective well-being and situational self-deception has not been investigated.

Moreover, while dispositional self-deception has been shown to have benefits in the long term due to its associations with subjective well-being and

personal adjustment (Dufner et al., 2019), some authors point out that in short-term self-deception might be associated with situational *negative affect* (Mele, 2000; Merchelbach & Merten, 2012; Sackeim & Gur, 1979). This negative affect might be felt before engaging in self-deception; however, after engaging in self-deception, there might be a decrease in negative affect or discomfort and an immediate increase in *positive affect* (Chance et al., 2011; 2015; Lauria et al., 2016). Robin and Beers (2001) showed that in college students, inflated self-perceptions were associated with increased positive affect immediately after task performance. However, in this study, in the long term, self-deceptive enhancers showed a decrease in subjective well-being. Furthermore, at the end of the academic year, individuals with inflated perceptions of their academic ability did not receive higher grades than did individuals with more realistic appraisals of their ability.

While self-deception and more specifically dispositional self-deception (self-enhancement) and increased subjective well-being have been studied more extensively (Dufner et al., 2019; Taylor et al., 2003), the relation with decreased anxiety and depression, the reduction of cognitive dissonance and reduced cognitive load have been addressed sporadically in only a handful of studies (Gur & Sackeim, 1979; Jian et al., 2019; Mijović-Prelec & Prelec, 2010; Sheridan et al., 2015).

Cognitive Discrepancy or Dissonance

Cognitive dissonance has been proposed by Festinger (1954) to represent the feeling of discomfort that arises when a person holds a different belief than their behaviour (e.g., cheating is wrong, but they are cheating on their romantic partner). This inconsistency between beliefs or attitudes and behaviour is very similar to the conceptualisation of self-deception. However, in cognitive dissonance, the emphasis is placed on the discomfort felt (e.g., guilt, shame), while in self-deception, the emphasis is on the way the situation is handled (e.g., changing the belief to excuse the behaviour). Thus, for example, if *cognitive dissonance* is present after engaging in unethical behaviour (e.g., cheating, lying; Merchelbach & Merten, 2012; Polage, 2018; Otgaar et al., 2023), this discomfort would be reduced by engaging in self-deception, leading to a decrease in cognitive dissonance and even a possible increase in positive affect immediately after this.

So far, the distinction between self-deception and cognitive dissonance has not been extensively investigated. Sackeim and Gur (1979) showed that self-deception is associated with an affective discrepancy measured using a physiologic response (galvanic skin conductance). To our knowledge, no other study measured cognitive dissonance in association with self-deception, either as situational or dispositional.

Cognitive Load

Cognitive load refers to the amount of information processed in our working memory at any moment. Interpersonal deception has a higher cognitive load as one has to remember both the truth and the details of their lies (Otgaar & Baker, 2018). It was proposed in the literature (Trivers, 2000; von Hippel & Trivers, 2011) that to reduce the cognitive load caused by deception, self-deception appeared as a byproduct of interpersonal deception with the final goal of deceiving others better. To our knowledge, only one empirical study investigated whether engaging in self-deception would lead to a decreased cognitive load and a more successful deception (Jian et al., 2019). In this study, the forward-looking paradigm was used to induce self-deception. They also included a deception group and used a scale to measure cognitive load (Chinese translated NASA Task Load Index, NASA-TLX; Xiao et al. 2005). Results showed that the cognitive load was statistically significantly higher for the deception group than the self-deception group, which was also compared with the control group (no manipulation, telling the truth), thus providing some initial support for this theory.

(Moral) Self-concept Maintenance

The idea of self-deceiving oneself to preserve one's (positive) moral self-concept maintenance has been extremely influential. Baumeister (1996), one of the first proponents of self-deception being motivated by the maintaining of well-being, argues that people are strongly motivated to "maintain favourable self-concepts of 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). From this perspective, self-deception is explained as the systematic, motivated avoidance of threatening or unpleasant information about the self, while elaborating and even constructing favourable feedback about the self.

When we try to break down this postulate, a few questions arise: what happens when individuals engage in self-deception as a way to maintain their self-concept (either positive or negative according to self-verification theory, Swann & Buhrmester, 2012)? Are they aware of the fact that they did something (e.g., "I cheated on a test"/"I did not cheat on a test") that is against their view of themselves (e.g., "I am an honest person"/"I am not an honest person")? And do they dismiss the information as not being vital for a bigger category (for example, cheating on a laboratory test would not be a defining characteristic of

being honest, Greve & Wentura, 2010)? This way of understanding the possible process of self-deception for the maintenance of moral self-concept also parallels the minimal criteria set by Gur & Sackeim (1979) of what constitutes self-deception: a) *the individual holds two contradictory beliefs simultaneously*: 1) "a person that cheats is not honest" ("I cheated on a test" is a behaviour, not a belief), 2) "I am an honest person"; b) *the individual is not aware of holding one of the beliefs*: not aware of holding the first belief: "a person who cheats is not an honest person"; c) *the individual's lack of awareness is motivated*: "if I become aware of the fact that I believe that a person who cheats is not honest, I will have to update my view on myself as to correspond with the behaviour I just engaged with"; instead, it is proposed that the person would choose different ways of resolving their inner conflict between the two beliefs such as engaging in cheating behaviour and not using it as a defining feature for their global view of themselves, to a certain extent (Greve & Wentura, 2010). Important to note is a similar construct: moral licensing or self-licensing in which an individual uses a good act ("I donated some money for charity today"), to cover up for something immoral or unethical ("I cheated on a test"). However, in self-concept maintenance and moral licensing, the lack of awareness or the genuine belief in a false belief is not a necessary condition, as it is in self-deception (Khalil & Feltovich, 2018).

Only a few studies tried to test the self-concept maintenance theory in relation to self-deception, by mostly looking at the way people react to receiving correcting information (e.g., feedback). For example, Baumeister & Cairns (1992) looked at *defensiveness* as measured by high social desirability and low anxiety, a measure of individual differences in repression which was seen as self-deceptive (see also Weinberger et al., 1979). They showed that participants high in defensiveness spent less time reading negative feedback about themselves than positive feedback compared with participants low on defensiveness. Moreover, these participants recalled more positive than negative words from the feedback received. Thus, it was concluded that highly socially desirable participants focused more on information that allowed them to keep their positive image of themselves. However, in this study, moral self-concept was not directly measured, nor was self-deception.

In another more recent, but similar study, which measured self-deception more directly, Peterson and colleagues (2003) looked at the behaviour of high or low self-deceivers (self-deception was measured using a validated self-deception scale, Self-Deceptive Enhancement, SDE; Paulhus, 1991). Participants were asked to play a gambling-type card game in which success was dependent on the ability to integrate information regarding task failure. They found that higher

trait self-deceivers choose to continue playing even when losing repeatedly, thus showing evidence of ignoring evidence of error, even when maladaptive (i.e., losing money in a game).

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.

Furthermore, while not directly measuring self-deception, but rather overclaiming, which could be confounded by other-deception, Mazar et al. (2008) investigated "the self-concept maintenance" theory in a series of experiments. This theory posits that individuals engage in dishonest acts (cheating on a test) to a limited extent, to gain some benefits (financial incentives), but not enough to update their moral self-concept and include their dishonesty in the update. In one of six experiments, Mazar et al. (2008) gave participants a test and offered half of the participants an opportunity to cheat. Afterwards, participants had to calculate their scores, communicate their final score to the investigators, and fill in a short "personality" questionnaire, which included two questions on morality (being an honest and moral person in contrast with the day before). The results showed that even if aware of overclaiming, participants in the cheating condition did not show evidence of considering their cheating behaviour when answering the two morality questions. The authors concluded that these results support the self-concept maintenance theory, suggesting that limited dishonesty "flies under the radar". This means that individuals do not update their self-concept in terms of honesty even though they might be aware that they overclaimed. However, many of the findings in this study failed to be replicated, and one major critique is that the dependent measure differed between the control condition and the cheating condition. More specifically, for the control condition, the number of correct answers was used as the outcome, while for the cheating condition, participants' self-report of their total score was used, which could explain at least partly the differences found between the two conditions (Verschuere et al., 2018).

Nonetheless, as the method used in this study is similar to the forward-thinking paradigm, one question arises whether people who cheat and integrate their cheating as an indication of their future performance (self-deceivers, thus not directly aware of their overclaiming) might also maintain their self-concept even when they engage in unethical behaviours such as cheating on a test.

Avoiding "Threatening" Information

While protecting oneself from the truth and/or threatening information is an influential motivation proposed by the psychanalytic authors for self-deception, these theories have not been directly addressed in empirical research. To our knowledge, no study looked directly at self-deception and traumatic events.

Researchers tested the hypothesis that self-deception arises or increases as a reaction to threatening information by investigating the reaction of participants to negative (social) feedback (e.g., Baumeister & Cains, 1992; Peterson et al., 2003; Sackeim & Gur, 1979). Initial studies showed that people mostly ignored negative feedback compared to positive feedback, thus avoiding possible threats to their self-view. However, in these studies, it was not measured if this avoidance of negative feedback was related to any motivation, and it could have been more of a general bias against negative feedback. To address this, Von Hippel et al. (2005) devised experiments in which they included what they called an "ego threat" task, measuring the reaction participants have to success and failure feedback. In these experiments, participants first received a task in which they could cheat in a test, which was used to measure rationalised cheating. Afterwards, in two other tasks, participants received both success and failure feedback, and they were later asked which of these tasks measured their important qualities better (the ego threat tasks). If participants rated the task at which they succeeded as more important, it would be explained as self-serving processing of information. The results showed that rationalised cheating was predicted by self-serving processing. However, in one of the experiments, they also measured trait self-deception, and it was not associated with the measure used in these experiments for rationalised cheating. Thus, while rationalised cheating was associated with a positive information processing bias, it was not associated with a measure of self-deception.

In two studies using the forward-thinking paradigm to induce and measure self-deception (Liu et al., 2025 and Mei et al., 2023), when presented with negative feedback, participants engaged in less self-deception (compared with ambiguous and positive feedback). Similarly, Johnson (1995) also showed that when exposed to negative feedback (e.g., failure), self-deception increased when the context was more ambiguous (e.g., an excuse for failure was available). Another study by Sloman et al. (2009) showed that the diagnosticity of the feedback did not matter (positive vs negative), yet self-deception increased when feedback was more vague (containing words which approximated such as "lower/higher than").

While these last findings suggest that self-deception might not be increased or elicited as a reaction to more negative feedback, it does not immediately reject the theory of individuals engaging in self-deception as a reaction to traumatic events. For example, the studies done on health-related information showed that when the diagnosticity would point to a bad diagnosis, participants would engage in more self-deceptive behaviours (e.g., keeping their hands more or less in cold water to align with their preferred diagnosis) (Kunda, 1987, 1990; Quattrone & Tversky, 1984). However, as with the case of the controversy around repression (Otgaar et al., 2019), which has also been defined as a defence mechanism in which information is stored in the unconscious, it could be that the mechanisms in place when traumatic events occur might be better elucidated by alternative science-based explanations such as encoding failure (see Dodier et al., 2024 for a discussion on repression and alternative science-based explanations).

CONCLUDING REMARKS

In this paper, we briefly presented some of the most influential philosophical perspectives on self-deception and the existing psychological theories. Furthermore, we put together an overview of operationalised self-deception in empirical psychology. Our main focus was to present a comprehensive overview of pioneering empirical research on motivations behind self-deception formation, going beyond theoretical speculations on this essentially latent phenomenon.

In psychological research, self-deception has been defined as either a defence mechanism (denial, repression, etc.) or as the inflation of one's self (abilities, traits, potential). Some authors postulate that self-deception has evolved as a byproduct of interpersonal deception, is similar to biases in information processing. Others see it as a type of motivationally-biased judgement. 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. While the question of self-deception being a case of motivated reasoning has been essential for a substantial part of the theoretical psychological literature, to the best of our knowledge, no study has yet provided a synopsis of empirical research addressing this.

The motivations proposed in the literature and covered in this review could be categorised into external and internal motivations. During the past decades, empirical studies were published investigating situational or trait/dispositional self-deception in relation to these motivations. The most consistent motivation

studied so far has been subjective *well-being*. Self-deception, particularly dispositional or trait self-enhancement, appears to have a generally positive association with subjective well-being and personal adjustment. Studies consistently show that individuals with higher levels of self-deception tend to report greater overall positive psychological traits (and negative associations with socially undesirable traits or characteristics such as anxiety or aggression). However, for other motivations such as increased *confidence*, an increase in short-term positive affect, a decrease in cognitive load, and a decrease in negative short-term affective states or discomfort (e.g., cognitive dissonance), only a few studies addressed them sporadically, showing for each proposed motivation some initial support. Future studies should address these latter motivations more methodically to further clarify their role.

Nonetheless, external motivations such as financial incentives, deceiving others, social recognition, and validation were more studied, with initial studies providing empirical support, especially for situational self-deception and in the short term. However, findings so far suggest that in the long-term these effects decrease, thus self-deception could be beneficial in the short-term, but costly in the long term. Future research could also focus on other motivations and long-term benefits or costs.

Another influential theory in self-deception is the motivation of maintaining one's (moral) self-concept. So far, we have found limited research showing that individuals with high self-deception tend to ignore more negative feedback and that dispositional self-deception is associated with measures of trait moral self-concept. Ignoring negative feedback might be more attributed to a systematic information processing bias. The association with moral self-concept could be explained by trait self-deception leading to inflated scores on the moral self-concept measure. Future studies could measure more distinctly the maintenance of the moral self-concept (e.g., before and after a task in which participants can be dishonest) and use paradigms which also measure situational self-deception and are more illustrative for applied settings. For example, in the legal field, false beliefs could have severe consequences such as miscarriages of justice. Thus, it is essential to have a better understanding of the implications different motivations could have on leading to false beliefs in the legal field.

Furthermore, while influential psychoanalytic theories propose self-deception as a protective mechanism against threatening information or trauma (see also work on repression), direct empirical research on this relationship is lacking. So far, some studies have shown that in health-related scenarios, self-deceptive behaviours increased when diagnoses were potentially negative. For example, studies have focused on situations such as receiving negative feedback as a way of measuring an individual's reaction to threatening information.

Despite some initial findings of higher trait self-deceivers ignoring negative feedback, in more recent studies, self-deception was shown to be increased in more ambiguous contexts or when feedback was vague, but not necessarily when this was negative. Future studies should clarify the role of context, ambiguity, and the nature of threatening information in triggering self-deceptive behaviours. Furthermore, a more direct investigation of the relationship between self-deception and traumatic events would bring even more clarity to this postulate.

All in all, while there were consistent results for many of the motivations, some inconsistencies in findings were also present. For example, in the case of self-deception leading to increased confidence, some studies showed increased confidence in situational self-deception immediately after self-deception, but not in the long term, while for trait self-deception, the association was with lower confidence when lying in the short term. These inconsistencies could be attributed to methodological differences, particularly in how self-deception is measured (situational vs. trait) and the specific contexts in which it was studied (e.g., college students, community vs. offenders). Future studies should aim to standardise measurements of self-deception and explore how different motivations influence different types or degrees of self-deception. Including measures of individual differences (e.g., negative personality traits such as Dark Traits) could further clarify these inconsistencies.

The present review represents a first attempt to provide a comprehensive overview of empirical data on motivated false beliefs. One main limitation of this review is that the literature search was not systematically done. For each motivation, multiple relevant databases (PubMed, World of Science, Science Direct, Scopus) were searched using specific keywords (e.g., "cognitive dissonance" AND ("self-deception" OR "self-enhancement"). Future research should focus on a systematic review of the literature on motivations and self-deception to ensure that essential empirical studies have not been missed in the process (such as studies not published in English, e.g., Wei et al., 2024). Additionally, for paradigms such as the forward-thinking paradigm used in multiple recent studies, a preliminary analysis of the effect sizes across them could be attempted.

These limitations notwithstanding, this review has direct implications for theory, methodology, and practice. While providing initial support for self-deception being influenced by various internal and external motivations, this review further emphasises the need for a critical perspective of the validity of the construct, both for situational and dispositional self-deception. Regarding its methodological implications, this review offers an overview of different paradigms and methods used so far in studying self-deception and emphasises the need for conducting studies using similar paradigms for better reproducibility and replicability of the research paradigms and better generalizability of the results.

In conclusion, this narrative review offers a comprehensive overview of the limited available empirical research on internal and external motivations proposed in the literature on self-deception. Our analysis revealed consistent support for subjective well-being as a primary motivation, while also highlighting the roles of external motivations such as financial incentives and social recognition. Although less explored, some limited research exists on additional motivations for self-deception, which could serve as a foundation for future investigations in this field. This construct still faces challenges in methodological consistency and the need for more robust empirical evidence in certain areas. Moving forward, several key directions for future research emerge, including the standardisation of measurement techniques, enhancing ecological validity, and considering interindividual differences. This review contributes to refining theoretical frameworks and informs practical applications in fields such as legal psychology.

Authors' note

We have no conflicts of interest to disclose. As this manuscript represents a narrative review, there is no data to support its findings.

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

Ashley, A., & Holtgraves, T. (2003). Repressors and memory: Effects of self-deception, impression management, and mood. *Journal of Research in Personality, 37*(4), 284–296. https://doi.org/10.1016/S0092-6566(02)00567-6

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., 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

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

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

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., 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

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

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

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

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

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. https://doi.org/10.3724/SP.J.1041.2022.00481

Fernbach, P. M., Hagmayer, Y., & Sloman, S. A. (2013). Effort denial in self-deception. *Organizational Behavior and Human Decision Processes*, *123*(1), 1-8. https://doi.org/10.1016/j.obhdp.2013.10.013

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*. https://doi.org/10.1177/001872675400700202

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.

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

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, behavioral, 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

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

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

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647. https://doi.org/10.1037/0022-3514.53.4.636

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

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

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

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

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., & Chang, L. (2014). Deceiving yourself to better deceive high-status compared to equal-status others. *Evolutionary psychology: an international journal of evolutionary approaches to psychology and behavior*, *12*(3), 635–654. https://doi.org/10.1177/147470491401200310

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

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

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

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., 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:*

A journal of the Association for Psychological Science, *14*(6), 1072–1095. https://doi.org/10.1177/1745691619862306

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

Pass, J. C. (2024, November 7). *The psychology of delusion: Understanding Don Quixote's quest.* Simply Put Psych. https://simplyputpsych.co.uk/monday-musings-1/the-psychology-of-delusion-understanding-don-quixotes-quest

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

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., & 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.

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

Polage, D. C. (2019). Liar, liar: Consistent lying decreases belief in the truth. *Applied Cognitive Psychology*, *33*(4), 527-536. https://doi.org/10.1002/acp.3489

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

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

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., 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

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

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

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

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

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., & 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

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

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

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

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. (1989). *Positive illusions: Creative self-deception and the healthy mind.* Basic Books/Hachette Book Group.

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.

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., Skowronski, J. J., Acar, O. A., Aczel, B., Bakos, B. E., Barbosa, F., Baskin, E., Bègue, L., Ben-Shakhar, G., Birt, A. R., Blatz, L., Charman, S. D., Claesen, A., Clay, S. L., . . . Yıldız, E. (2018). Registered Replication Report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.1177/2515245918781032

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/

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

von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *The Behavioral and Brain Sciences*, *34*(1), 1–56. https://doi.org/10.1017/S0140525X10001354

Wei, F., Ying, Y., Yanqiu, D., & Wenjie, Z. (2024). The Effects of Competitive Situations and Reward Motivation on Self-Deception: Evidence From Eye Movements[J]. *Journal of Psychological Science, 47(2):* 411-423. **范**伟, 杨颖, **董**艳秋, 张文洁. (2024). 竞争情境下奖赏动机对自我欺骗的影响：眼动证据*[J]. ***心理科学****, 47(2)*: 411-423.

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, 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

Xiao YM, Wang ZM, Wang MZ, Lan YJ. (2005). [The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index]. Zhonghua lao Dong wei Sheng zhi ye Bing za zhi = *Zhonghua Laodong Weisheng Zhiyebing Zazhi = Chinese Journal of Industrial Hygiene and Occupational Diseases. Jun; 23(3):*178-181.

Yang, Y., Zhong, B., Zhang, W., & Fan, W. (2024). The Effect of Comparative Direction and Comparative Gap on Self-Deception. *Psychology research and behavior management*, *17*, 2819–2834. https://doi.org/10.2147/PRBM.S467437