The Effects of Cognitive Ambiguity and Motivational Intensity on Conscious and Unconscious Self-Deception

Authors: Zhong Luojin, Ru Taotao, Fan Meng, Morey, Morey

Date: 2019-09-22T00:00:00+00:00

Abstract

Three experiments were conducted to investigate the existence of unconscious and conscious self-deception, and the influence of cognitive ambiguity level and motivational intensity on both forms. Experiment 1 demonstrated that self-deception exists in two distinct forms: unconscious and conscious; Experiment 2 revealed that as cognitive ambiguity decreased, the incidence of unconscious self-deception declined significantly, whereas the incidence of conscious self-deception increased significantly; Experiment 3 showed that with increased motivational intensity, conscious self-deception increased significantly, while unconscious self-deception remained unchanged. These findings indicate that self-deception is a genuine phenomenon comprising two distinct forms, each characterized by different underlying mechanisms.

Full Text

The Effects of Cognitive Vagueness and Motivation on Conscious and Unconscious Self-Deception

ZHONG Luojin[1], RU Taotao[2], FAN Meng[3], MO Lei

[1]Institute of Basic Education Training and Research, South China Normal University, Guangzhou 510631, China
[2]Lab of Light and Physio-psychological Health, National Center for International Research on Green Optoelectronics, South China Normal University, Guangzhou 510006, China
[3]Student Career Service Center, Guangdong University of Technology, Guangzhou 510006, China
School of Psychology, South China Normal University, Guangzhou 510631, China

Abstract

Three experiments investigated the existence of unconscious and conscious self-deception and examined how cognitive vagueness and motivation intensity influence these two forms of self-deception. Experiment 1 demonstrated that both unconscious and conscious self-deception exist. Experiment 2 found that as cognitive vagueness decreased, the number of unconscious self-deceivers declined significantly while the number of conscious self-deceivers increased substantially. Experiment 3 revealed that increased motivation intensity led to more conscious self-deceivers but did not significantly affect unconscious self-deceivers. These findings suggest that self-deception is a genuine phenomenon with two distinct forms governed by different underlying mechanisms.

Keywords: self-deception; consciousness; motivation; cognitive vagueness

1. Introduction

Self-deception refers to the phenomenon where an individual comes to believe a proposition that contradicts a belief they already hold (Demos, 1960). Although self-deception appears ubiquitous in daily life, its scientific investigation presents formidable challenges. The deceiver and the deceived are one and the same, creating two logical paradoxes: First, the process (dynamic) paradox—how can an individual deceive themselves? That is, what mechanisms enable a person to simultaneously believe a proposition and its opposite? Second, the outcome (static) paradox—how can an individual maintain two contradictory beliefs in psychological equilibrium? Assuming a person enters a state of self-deception, how do they sustain this psychologically imbalanced state of endorsing both a positive and a negative belief (Mele, 1997)?

Due to these unique characteristics, the very existence of self-deception has remained a central focus of debate in the field. Philosophers have extensively theorized about this issue (McKay & Dennett, 2009), while psychologists have approached it empirically (Johnson & Fowler, 2011; Kurzban, 2012; Westland & Shinebourne, 2009; Zhu, Xi, & Wu, 2016). To date, psychological research has yielded important findings in three main areas.

First, studies using rigorous experimental controls have documented self-deceptive behaviors, providing direct evidence for its existence. For instance, Gur and Sackeim (1979) found that participants would self-deceptively deny recognizing their own voices after experiencing failure. Quattrone and Tversky (1984) induced longevity desires through instructions (the longer one endured cold water, the healthier their heart would be and the longer they would live), resulting in participants significantly increasing their cold-water endurance on a second trial while denying any influence from the instructions. This experimental paradigm has been replicated in subsequent research (Fernbach, Hagmayer, & Sloman, 2014). Whitson and Galinsky (2008) found that participants experiencing helplessness from failing an impossible task subsequently perceived more meaningful patterns in random snowflake distributions,

thereby increasing their sense of control. Gebauer, Göritz, Hofmann, and Sedikides (2012) asked participants to write about their most beloved person and then administered both explicit and implicit tests. Results showed that participants explicitly reported loving this person more than themselves, yet implicitly demonstrated self-love exceeding love for others. Researchers have also identified self-deception across numerous domains, where individuals deceive themselves into believing they are more moral, attractive, or competent than they actually are (Epley & Whitchurch, 2008; Gilovich, 2008; Lönnqvist, Irlenbusch, & Walkowitz, 2014; Mijović-Prelec & Prelec, 2010).

Second, researchers have investigated the conditions and influencing factors of self-deception. Mele (2010) proposed four necessary and sufficient conditions for self-deception: (1) the individual desires a false belief to be true; (2) the individual processes data related to this false belief with bias; (3) this biased data supports the false belief; and (4) the individual possesses more data supporting the false belief. Michel and Newen (2010) argued that self-deceptive behavior first requires self-deceptive motivation. Sloman, Fernbach, and Hagmayer (2010) found that beyond motivation, self-deception also requires cognitive vagueness. Other studies have examined influencing factors, such as Ren et al. (2018), who found that individuals with strong self-control can better avoid the disadvantages of self-deception, while high-status individuals can inhibit self-deception in low-status individuals.

Third, researchers have examined the utility of self-deception. On one hand, self-deception can benefit individuals by enhancing confidence and happiness, yielding greater benefits, and avoiding or reducing punishment. Erez, Johnson, and Judge (1995) noted that self-deception can improve subjective well-being. Surbey and McNally (1997) found that individuals prone to self-deception exhibit greater cooperativeness. Lopez and Fuxjager (2012) argued that self-deceptive individuals view themselves more positively, enhancing their social image and securing higher social status and greater social benefits. Sheridan, Boman, Mergler, and Furlong (2015) found significant correlations between college students' life satisfaction and self-deception. On the other hand, because self-deception sacrifices informational authenticity, it can also incur costs (Von Hippel & Trivers, 2011). Chance, Norton, Gino, and Ariely (2011) found that self-deception remained stable even when facing potential losses. In their study, both experimental and control groups completed two tests, receiving payment for each correct answer. The experimental group had answer keys printed at the bottom of the first test, while the control group did not. After completing the first test, both groups proceeded to the second test without answer keys. Before beginning the second test, participants were asked to accurately predict their performance.

Results showed that the experimental group predicted significantly higher scores on the second test than the control group, indicating that they self-deceptively attributed their cheating performance on the first test to personal ability. Further investigation revealed that even when substantial rewards were offered for

prediction accuracy, the experimental group continued to predict they would achieve the same high scores as on the first test. This persistent self-deception actually impaired their prediction accuracy and prevented them from obtaining greater rewards.

A review of empirical research on self-deception reveals that despite attracting considerable attention and yielding rich findings, the field' s focus has seemingly shifted from the existence question to mechanisms of emergence and evolution (Fan, Yang, Liu, & Fu, 2017). However, breakthroughs in understanding self-deception mechanisms have remained elusive, likely because the fundamental question of its existence has not been fully resolved. Analysis of previous research reveals that nearly all studies have overlooked a crucial issue: whether self-deceivers truly deceive themselves. For example, Chance et al. (2011) found that participants self-deceptively attributed answer-copying performance to personal ability, yet it remains unclear whether all experimental participants genuinely believed this.

The question of self-deception' s intentionality has long been debated in philosophy (Ma & Chen, 2016; Neuber, 2016; Lynch, 2016). Intentionalism posits that people deliberately deceive themselves, resulting in two contradictory beliefs in consciousness and thus both static and dynamic paradoxes (Davidson, 2004). Non-intentionalism argues that people do not intentionally deceive themselves but rather come to believe the opposite proposition through biased cognitive processing, resulting in only one biased belief in consciousness and neither paradox (Mele, 2010). We propose that according to intentionalism, although individuals engage in self-deceptive behavior, they cannot truly deceive themselves because two contradictory beliefs coexist in consciousness. According to non-intentionalism, individuals not only deceive themselves but likely do so completely.

Based on this analysis, we hypothesize that previous research may have conflated two forms of self-deception: conscious and unconscious. Conscious self-deception refers to individuals who engage in self-deceptive behavior while being aware of it; such individuals cannot truly deceive themselves because they harbor two contradictory beliefs, making this form situational. Unconscious self-deception refers to individuals who engage in self-deceptive behavior without awareness; these individuals possess only one belief and can truly deceive themselves, making this form stable. Both perspectives have received empirical support. For instance, intentionalism has been supported in interpersonal self-deception research, where individuals deliberately self-deceive for personal gain (Lu & Chang, 2014; Smith, Trivers, & Von Hippel, 2017; Seiffert-Brockmann & Thummes, 2017). Non-intentionalism has received support from studies on self-enhancement phenomena (Leslie & Mckay, 2016). However, no previous experiments have directly tested whether these represent conscious versus unconscious self-deception. Moreover, existing experiments have used different operational definitions to investigate different psychological phenomena. Investigating whether two forms of self-deception exist for the same phenomenon

under a unified operational definition would help resolve these controversies.

As previously noted, Chance et al.' s (2011) study likely involved both forms of self-deception. Their use of a forward-looking paradigm differs from previous retrospective paradigms that identified self-deception through inconsistencies between behavioral responses and self-reports. This new method holds considerable research promise (Chance, Gino, Norton, & Ariely, 2015). However, the forward-looking paradigm cannot currently distinguish between conscious and unconscious self-deception.

Therefore, we modified Chance et al.' s (2011) method to test whether two types of self-deception can be identified within a single experiment. The new method added a third test to the original two, requiring participants to predict their performance on an equivalent third test form. Thus, experimental group participants had two test scores before predicting their performance on Test 3: one from the first test with answer keys and one from the second test without answers. If participants' predictions for Test 3 approximated their actual performance on Test 2 (without answers), this would indicate conscious self-deception during their Test 2 predictions, as they adjusted their self-assessment after seeing their true ability. If predictions for Test 3 remained close to their performance on Test 1 (with answers), this would indicate unconscious self-deception during Test 2 predictions, as participants persisted in attributing their cheating performance to personal ability despite knowing their true level. If predictions for Test 3 fell between Test 1 and Test 2 scores, this would suggest both forms coexisted: conscious self-deceivers would shift to using their true score (Test 2) as a reference, yielding predictions far below Test 1 scores, while unconscious self-deceivers would maintain their deception, yielding predictions far above Test 2 scores. In this third scenario, we could classify each participant as an unconscious self-deceiver (if their Test 3 prediction was closer to Test 1) or a conscious self-deceiver (if closer to Test 2). Experiment 1 employed this new method to investigate the existence of both forms of self-deception.

2. Experiment 1

2.1 Purpose

This experiment examined whether unconscious and conscious self-deception exist using the new experimental method.

2.2 Method

Participants. A total of 165 college students participated, with 61 males, ranging in age from 18 to 24 years (M = 20.47). The experimental group comprised 103 participants, and the control group comprised 62. Because some experimental group participants might not view the answers, which would violate experimental requirements, the experimental group was intentionally larger. All participants received compensation based on their performance.

Materials. Three test forms were used, each containing ten general knowledge questions covering history, literature, geography, and sports. To ensure participants perceived the forms as equivalent, we conducted pilot testing. First, 100 general knowledge questions were administered to 100 pilot participants. Questions were then categorized by difficulty based on accuracy rates. Each test form contained 4 difficult, 4 moderate, and 3 easy questions, with difficulty distributed across forms using a Latin square design (123, 231, 312). To verify that the forms would not be suspected as non-equivalent, 60 participants were randomly assigned to complete one form. An ANOVA on test scores revealed no significant differences among forms ($M = 4.30 \pm 1.45$, $M = 5.20 \pm 1.15$, $M = 4.40 \pm 2.18$, $F(2, 59) = 1.87$, $p > 0.05$), confirming that the forms were not readily distinguishable in terms of difficulty or content.

Design. A single-factor between-subjects design was employed with two levels: whether answer keys were provided below Test 1. In the experimental group, Test 1 had answer keys printed clearly at the bottom, while Tests 2 and 3 contained only questions. In the control group, all three tests lacked answer keys, with all other aspects identical to the experimental group.

Procedure. Participants completed tests individually in a quiet classroom, seated one meter apart to prevent communication or answer copying. The experimenter informed participants they would complete three similar test forms, receiving ¥0.50 for each correct answer. All participants first completed Test 1, which was then graded by the experimenter, who provided total earnings. Before Test 2, the experimenter emphasized its equivalence to Test 1, asked participants to quickly review the questions (ensuring experimental group participants recognized the absence of answer keys), and then requested predictions of how many questions they would answer correctly. After completing Test 2, participants received their actual scores and earnings. Finally, participants received Test 3. Upon distribution, the experimenter again emphasized equivalence to previous tests and asked participants to quickly review questions before predicting their performance. Participants were not informed that Test 3 would not be administered after prediction. Test completion time was 3 minutes; prediction time was 30 seconds.

2.3 Results and Discussion

Data from all 165 participants were analyzed. First, two participants (both from the experimental group) were excluded for making careless predictions (predicting 0 or 1 correct answers). Second, honest participants in the experimental group who did not view answers were removed. Since pilot testing revealed average scores above 4, participants scoring below 4 on Test 1 likely did not cheat. Twenty-six such participants were excluded. Additionally, eight participants whose Test 2 scores exceeded their Test 1 scores were eliminated, as equivalent forms made this pattern indicative of non-cheating on Test 1. The final sample consisted of 67 experimental group participants and 62 control group participants.

Table 1 presents descriptive statistics for both groups. Independent samples t-tests using Cohen' s formula calculated effect size d, with paired samples t-tests using Morris and DeShon' s formula (d = 0.2, 0.5, and 0.8 representing small, medium, and large effects, respectively). Test 1 scores differed significantly between groups, with the experimental group scoring far higher than the control group ($t(127) = 9.27$, $p < 0.001$, $d = 1.64$), confirming that experimental group participants viewed the answers. Analysis of Test 2 revealed no significant difference in actual performance between groups ($t(127) = 1.50$, $p > 0.05$), indicating equivalent ability levels. However, predicted scores for Test 2 differed significantly, with the experimental group predicting higher performance than the control group ($t(127) = 3.43$, $p < 0.001$, $d = 0.60$). This suggests experimental participants self-deceptively attributed their cheating performance to personal ability, expecting equivalent performance on Test 2. This pattern replicates Chance et al. (2011).

To identify unconscious and conscious self-deceivers, we compared experimental group participants' Test 3 predictions with their Test 1 (cheating) and Test 2 (actual) scores. Test 3 predictions were significantly higher than Test 2 scores ($t(66) = 3.34$, $p < 0.001$, $d = 0.41$) and significantly lower than Test 1 scores ($t(66) = 10.77$, $p < 0.001$, $d = 1.32$), indicating that predictions fell between cheating and actual performance. As hypothesized, this suggests both unconscious and conscious self-deceivers were present.

To classify individual participants, we compared each participant' s Test 3 prediction to their Test 1 and Test 2 scores. Participants whose Test 3 predictions approximated their Test 1 scores were classified as unconscious self-deceivers; those approximating Test 2 scores were classified as conscious self-deceivers. Nine participants whose predictions fell exactly midway between scores were excluded. The remaining 58 participants comprised 16 unconscious self-deceivers and 42 conscious self-deceivers (see Table 2).

Conscious self-deceivers showed declining predictions from Test 1 cheating scores to Test 2 predictions, with Test 3 predictions approaching actual ability. Unconscious self-deceivers maintained predictions close to their initial cheating scores across Tests 2 and 3. This pattern supports our hypotheses and explains why Chance et al. (2011) found that experimental participants predicted higher scores than controls even when rewarded for accuracy: unconscious self-deceivers failed to adjust their predictions, maintaining them near their cheating performance.

Were conscious self-deceivers truly self-deceiving? Comparing their Test 2 predictions to actual performance revealed significantly higher predictions ($t(41) = 2.85$, $p < 0.01$, $d = 0.47$), confirming self-deceptive behavior. Notably, conscious self-deceivers scored significantly higher on Test 1 than unconscious self-deceivers ($t(56) = 4.01$, $p < 0.001$, $d = 1.26$), yet their actual ability (Test 2 performance) did not differ ($t(56) = 0.18$, $p > 0.05$). This indicates conscious self-deceivers viewed more answers, making their cheating behavior too salient to deny unconsciously. Unconscious self-deceivers viewed fewer answers, leaving

cognitive ambiguity about whether cheating occurred. This aligns with Sloman et al.' s (2010) finding that self-deception requires vagueness. However, Experiment 1' s conclusion about cognitive vagueness remains speculative, as different methods were used. Experiment 2 directly manipulated cognitive vagueness to examine its specific effects on both forms of self-deception.

3. Experiment 2

3.1 Purpose

This experiment investigated how cognitive vagueness influences unconscious and conscious self-deception.

3.2 Method

Design. A single-factor between-subjects design manipulated cognitive vagueness (high vs. low). Cognitive vagueness refers to the degree to which participants consciously detected their cheating behavior; clearly knowing one had viewed answers represents low vagueness, while uncertainty represents high vagueness. Vagueness was manipulated through answer presentation format. In the low vagueness condition, Test 1 answers appeared upside-down at the bottom of the page, requiring effort to read and thus making cheating behavior more salient. In the high vagueness condition, answers appeared right-side-up as in Experiment 1, making them easy to view.

Manipulation check. To validate this manipulation, 60 participants were randomly assigned to upright or inverted answer conditions and completed the same procedure as Experiment 1. Afterward, they completed an anonymous questionnaire with two items: (1) "Did you notice the answers below the questions?" (yes/no), and (2) "To what extent did the answers help you solve the problems?" (0 = no help to 5 = very helpful). Participants completed this unsupervised and deposited it in a designated box.

All 60 participants completed the tests and questionnaire. Test scores did not differ between conditions ($t(58) = 0.31$, $p = 0.75$; $M\_upright = 7.36 \pm 1.63$; $M\_inverted = 7.50 \pm 1.68$), indicating both groups cheated. All participants noticed the answers. Although most were unwilling to admit the answers helped (responses below 1), the upright condition admitted significantly less help ($t(58) = 2.98$, $p < 0.01$; $M\_upright = 0.47 \pm 0.63$; $M\_inverted = 0.97 \pm 0.67$). This indicates inverted-answer participants had clearer awareness of their cheating behavior, representing lower cognitive vagueness. Thus, the manipulation was effective.

Procedure. Experiment 2 used the same group testing format and identical procedures as Experiment 1, with the same compensation scheme.

3.3 Results and Discussion

Data analysis followed Experiment 1' s procedure. First, three participants (two from the high vagueness condition) were excluded for careless predictions. Second, 29 non-cheating participants were removed (15 from high vagueness). The final sample included 33 participants in the high vagueness condition and 35 in the low vagueness condition.

Test 1 scores did not differ between vagueness conditions ($t(66) = 0.66$, $p > 0.05$), indicating the manipulation did not affect cheating behavior. Within-group analyses revealed that in the low vagueness condition, Test 3 predictions were marginally higher than Test 2 scores ($t(34) = 1.94$, $p = 0.06$, $d = 0.36$) and significantly lower than Test 1 scores ($t(34) = 11.04$, $p < 0.001$, $d = 1.88$). Similarly, in the high vagueness condition, Test 3 predictions were significantly higher than Test 2 scores ($t(32) = 3.87$, $p < 0.001$, $d = 0.68$) and significantly lower than Test 1 scores ($t(32) = 6.81$, $p < 0.001$, $d = 1.10$). This pattern replicates Experiment 1, indicating both forms of self-deception existed regardless of vagueness level.

Table 3 presents results for both conditions. To identify self-deceiver types, we used the same classification method as Experiment 1, excluding participants whose predictions fell midway between scores (4 in low vagueness, 3 in high vagueness). Chi-square tests with effect size (w) showed significant differences in unconscious self-deceivers between conditions ($^2(1) = 6.35$, $p < 0.05$, $w = 0.26$), with more unconscious self-deceivers in the high vagueness condition. Conscious self-deceivers also differed significantly ($^2(1) = 4.03$, $p < 0.05$, $w = 0.21$), with fewer in the high vagueness condition. Thus, reduced cognitive vagueness decreased unconscious self-deception while increasing conscious self-deception.

These results suggest that when cheating requires greater effort, cognitive clarity increases, making it harder to unconsciously deceive oneself and reducing unconscious self-deception. This aligns with Sloman et al.' s (2010) finding that vagueness affects unconscious self-deception. However, Experiments 1 and 2 used fixed rewards, creating relatively constant motivation. Experiment 3 manipulated reward magnitude to examine how motivation influences both forms of self-deception.

4. Experiment 3

4.1 Purpose

This experiment examined how motivation affects conscious and unconscious self-deception.

4.2 Method

Participants. One hundred twenty college students (61 males) aged 17-25 years ($M = 20.52$) participated, with 60 assigned to each motivation condition.

Design. A single-factor between-subjects design manipulated motivation level (high vs. low). Both conditions had upright answer keys on Test 1. High motivation offered ¥1.00 per correct answer; low motivation offered ¥0.50. All participants received compensation based on performance.

Procedure. Experiment 3 used the same group testing format and identical procedures as Experiment 2.

4.3 Results and Discussion

Data analysis followed previous procedures. No participants made careless predictions (0 or 1). Twenty non-cheating participants were removed from the low motivation condition and six from the high motivation condition, leaving 54 cheaters in the high motivation condition and 40 in the low motivation condition.

A chi-square test on cheating rates confirmed the motivation manipulation was effective ($^{2}(1) = 9.62$, $p < 0.01$, $w = 0.28$), with more participants cheating in the high motivation condition. Table 5 presents performance across conditions, with no significant differences in predictions or actual scores ($p > 0.05$).

In both conditions, Test 3 predictions were significantly higher than Test 2 scores (high: $t(53) = 2.77$, $p < 0.01$, $d = 0.39$; low: $t(39) = 3.71$, $p < 0.001$, $d = 0.59$) and significantly lower than Test 1 scores (high: $t(53) = 10.63$, $p < 0.001$, $d = 1.44$; low: $t(39) = 7.10$, $p < 0.001$, $d = 1.13$), replicating previous patterns.

Classification of self-deceiver types (excluding 2 low-motivation and 4 high-motivation midpoint participants) revealed no significant difference in unconscious self-deceivers between conditions (10 vs. 14; $^{2}(1) = 0.83$, $p > 0.05$). However, conscious self-deceivers differed significantly (40 vs. 24; $^{2}(1) = 8.57$, $p < 0.01$, $w = 0.27$; see Table 6). Thus, motivation increased conscious self-deception but did not affect unconscious self-deception, further supporting their distinct nature.

These findings suggest that when stronger deception motivation exists (i.e., when individuals hope to gain more through self-deception), they are more likely to be conscious self-deceivers. This aligns with Von Hippel and Trivers' s (2011) view of self-deception as an evolved, sophisticated deception strategy. Lu (2012) termed this interpersonal self-deception, which, due to its strong motivational component, is likely conscious. Individuals cannot unconsciously deceive themselves but instead employ a more effective, reliable deception strategy to maximize gains. Consequently, conscious self-deception resembles interpersonal deception, being more situational and controllable.

5. General Discussion

Three experiments using a new forward-looking paradigm examined the existence of unconscious and conscious self-deception and how cognitive vagueness and motivation influence each form. Experiment 1 identified both conscious and

unconscious self-deception. Experiment 2 demonstrated that reduced cognitive vagueness decreased unconscious self-deceivers while increasing conscious self-deceivers. Experiment 3 showed that increased motivation enhanced conscious self-deception without affecting unconscious self-deception.

5.1 Characteristics of Unconscious and Conscious Self-Deception

The three experiments reveal distinct characteristics of the two forms. In Experiment 1, unconscious self-deceivers maintained predictions close to their cheating scores across Tests 2 and 3, whereas conscious self-deceivers initially self-deceived on Test 2 but subsequently adjusted predictions to their true ability on Test 3. This suggests unconscious self-deceivers truly deceive themselves from the outset, stably believing their ability matches their cheating performance. Conscious self-deceivers, aware of their deception, adjust predictions situationally. Experiments 2 and 3 further demonstrated differential effects of cognitive vagueness and motivation, confirming distinct characteristics.

Overall, unconscious self-deception is stable, truly deceives the individual at the conscious level, resembles honest behavior, and is primarily influenced by cognitive vagueness. Conscious self-deception is situational, changes with context, does not truly deceive at the conscious level, resembles interpersonal deception, and is susceptible to various factors including motivation, cognitive vagueness, time, self-control, and social status (Chance, Gino, Norton, & Ariely, 2015; Chance & Norton, 2015; Ren et al., 2018).

5.2 The Necessity of Distinguishing the Two Forms

Classifying unconscious versus conscious self-deception is valuable not only because they exhibit different characteristics but also because it has theoretical significance. First, it allows scientific differentiation of previous findings based on whether self-deceivers are aware of their deception, enhancing understanding of prior research. For example, Epley and Whitchurch' s (2008) finding that participants self-deceptively believed they were more attractive than reality likely reflects unconscious self-deception, as the experiment created substantial cognitive vagueness about whether enhanced photos truly represented oneself. Conversely, Lu' s (2012) interpersonal self-deception likely represents conscious self-deception, as its purpose is to deceive others for personal gain.

Second, this distinction may illuminate underlying mechanisms. Because conscious self-deception involves contradictory beliefs, it likely requires more cognitive resources and control, implicating greater prefrontal cortex involvement. Unconscious self-deception, resembling honest behavior, may proceed more automatically with less cognitive control. This speculation is supported by neuroimaging studies showing that self-deception activates frontal and parietal regions more strongly (Farrow, Burgess, Wilkinson, & Hunter, 2015; Sharma, Modi, Khushu, & Mandal, 2010), likely reflecting the cognitive control demands of conscious self-deception.

Third, this distinction may resolve long-standing theoretical controversies, such as the intentionalism versus non-intentionalism debate, which may simply reflect two different forms of self-deception. Fourth, it addresses the two paradoxes: static paradox disappears in unconscious self-deception because only one belief exists, while in conscious self-deception, the paradox is resolved through increased cognitive resources, much like interpersonal deception.

Indeed, most current research cannot effectively distinguish self-deception types, limiting mechanistic understanding. As Trivers (2011) noted, despite over half a century of research, the field remains in its infancy without breakthrough progress, largely because researchers have conflated different types with different mechanisms. Future research that effectively distinguishes unconscious and conscious self-deception will greatly advance understanding of its internal mechanisms.

5.3 The Relationship Between Unconscious and Conscious Self-Deception

Although distinct, the two forms are not unrelated. Experiment 2 showed that unconscious self-deception increased with cognitive vagueness while conscious self-deception decreased, suggesting they may transform into one another. Chance et al. (2015) also found that self-deception diminishes over time with repeated exposure to one' s true self.

However, Experiment 3 showed that increased external motivation enhanced conscious but not unconscious self-deception, suggesting they may not interconvert. This may reflect insufficient motivation strength differences. The experiment found more cheaters in the high motivation condition but no difference in the number of answers copied (mean scores did not differ). The larger cheating population provided a greater base for conscious self-deception. Stronger motivation that prompted copying more answers might increase conscious awareness of cheating, potentially shifting individuals from unconscious to conscious self-deception.

Furthermore, Experiment 2' s results suggest a more speculative hypothesis: conscious and unconscious self-deception may represent two points on a continuous gradient. Individuals may initially engage in unconscious self-deception, shifting to conscious self-deception when conditions demand it. From a cognitive economy perspective, unconscious self-deception consumes fewer resources than conscious self-deception. This hypothesis requires future experimental verification.

5.4 Significance and Future Directions

This study proposed and validated the existence of two self-deception forms and clarified their properties. Unlike previous research that identified self-deception at the group level through comparisons (e.g., Chance et al., 2011; Sloman et al., 2010), our method attempts to identify self-deception at the individual level

using delayed inference. Although this represents an initial step, it moves toward individual-level analysis.

Future research has a long path to elucidate self-deception mechanisms. Our classification will facilitate this endeavor. For example, cognitive experiments could test whether unconscious self-deceivers, having no internal conflict, show response times for self-deceptive responses matching correct responses, whereas conscious self-deceivers, experiencing conflict, show longer response times. Additionally, investigating the boundary conditions and mechanisms of conversion between forms could help people harness both types for beneficial purposes.

This study also highlights the complexity of self-deception's mechanisms. Different types and domains may involve entirely different processes, requiring careful differentiation in future research. Future studies should improve paradigms to identify self-deception in real-time at the individual level and employ advanced cognitive neuroscience methods to investigate mechanisms underlying different self-deception types.

In conclusion, three experiments demonstrate that self-deception exists in two forms: conscious and unconscious. Conscious self-deception increases with reduced cognitive vagueness and enhanced motivation, while unconscious self-deception decreases with reduced vagueness but remains unaffected by motivation.

References

Chance, Z., Gino, F., Norton, M. I., & Ariely, D. (2015). The slow decay and quick revival of self-deception. *Frontiers in Psychology*, 6, 1075.

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 of the United States of America*, 108, 15655–15659.

Chance, Z., & Norton, M. I. (2015). The what and why of self-deception. *Current Opinion in Psychology*, 6, 104-107.

Davidson, D. (2004). *Problems of rationality*. New York: Oxford University Press.

Demos, R. (1960). Lying to oneself. *The Journal of Philosophy*, 57(18), 588–595.

Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self-recognition. *Personality and Social Psychology Bulletin*, 34(9), 1159–1170.

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.

Farrow, T. F., Burgess, J., Wilkinson, I. D., & Hunter, M. D. (2015). Neural correlates of self-deception and impression-management. *Neuropsychologia*, 67, 159-174.

Fan, W., Yang, B., Liu, J., & Fu, X. L. (2017). Self-deception: For adjusting individual psychological states. *Advances in Psychological Science*, 25(8), 1349-1359. [范伟, 杨博, 刘娟, 傅小兰. (2017). 自我欺骗：为了调节个体心理状态. 心理科学进展, 25(8), 1349-1359.]

Fernbach, P. M., Hagmayer, Y., & Sloman, S. A. (2014). Effort denial in self-deception. *Organizational Behavior and Human Decision Processes*, 123(1), 1-8.

Gebauer, J. E., Göritz, A. S., Hofmann, W., & Sedikides, C. (2012). Self-love or other-love? Explicit other-preference but implicit self-preference. *PLoS One*, 7(7), e41789.

Gilovich, T. (2008). *How we know what isn' t so.* New York: Free Press.

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.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, 477(7364), 317-320.

Kurzban, R. (2012). On the advantages of being wrong. *BioScience*, 62(5), 516-517.

Lu, H.J. (2012). Self-deception: Deceiving yourself to better deceive others. *Acta Psychologica Sinica*, 44(9), 1265-1278. [陆慧菁. (2012). 自我欺骗：通过欺骗自己更好地欺骗他人. 心理学报, 44(9), 1265-1278.]

Lu, H. J., & Chang, L. (2014). Deceiving yourself to better deceive high-status compared to equal-status others. *Evolutionary Psychology*, 12(3), 635-654.

Lopez, J. K., & Fuxjager, M. J. (2012). Self-deception' s adaptive value: Effects of positive thinking and the winner effect. *Consciousness and Cognition*, 21(1), 315-324.

Lönnqvist, J.-E., Irlenbusch, B., & Walkowitz, G. (2014). Moral hypocrisy: Impression management or self-deception? *Journal of Experimental Social Psychology*, 55, 53-62.

Leslie, V. D. L., & Mckay, R. (2016). The optimist within? Selective sampling and self-deception. *Consciousness and Cognition*, S1053810016300770.

Lynch, K. (2016). Willful ignorance and self-deception. *Philosophical Studies*, 173(2), 505-523.

Ma, H.L., & Chen, G. (2012). On self-deception: Intentionalism vs. non-intentionalism. *Journal of Dialectics of Nature*, 38(5), 49-55. [马翰林, 陈刚. (2016). 自我欺骗问题研究——意向论 vs. 非意向论. 自然辩证法通讯, 38(5), 49-55.]

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6), 493–561.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20(1), 91–102.

Mele, A. R. (2010). Approaching self-deception: How Robert Audi and I part company. *Consciousness and Cognition*, 19(3), 745–750.

Michel, C., & Newen, A. (2010). Self-deception as pseudo-rational regulation of belief. *Consciousness and Cognition*, 19(3), 731–744.

Mijović-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: A model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538), 227–240.

Neuber, S. (2016). Self-awareness and self-deception: A sartrean perspective. *Continental Philosophy Review*, 49(4), 1–23.

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.

Ren, M. M., Zhong, B. W., Fan, W., Dai, H. M., Yang, B., Zhang, W. J., Yin, Z. X., Liu, J., Li, J., & Zhan, Y. L. (2018). The influence of self-control and social status on self-deception. *Frontiers in Psychology*, 9, 1256.

Seiffert-Brockmann, J., & Thummes, K. (2017). Self-deception in public relations: A psychological and sociological approach to the challenge of conflicting expectations. *Public Relations Review*, 43(1), 133–144.

Sharma, M., Modi, S., Khushu, S., & Mandal, M. K. (2010). Neural activation pattern in self-deceivers. *Psychological Studies*, 55(1), 71–76.

Sheridan, Z., Boman, P., Mergler, A., & Furlong, M. J. (2015). Examining well-being, anxiety, and self-deception in university students. *Cogent Psychology*, 2(1), 993850.

Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, 115(2), 268–281.

Smith, M. K., Trivers, R., & Von Hippel, W. (2017). Self-deception facilitates interpersonal persuasion. *Journal of Economic Psychology*, S0167487016301854.

Surbey, M. K., & McNally, J. J. (1997). Self-deception as a mediator of cooperation and defection in varying social contexts described in the iterated prisoner' s dilemma. *Evolution and Human Behavior*, 18(6), 417–435.

Trivers, R. (2011). *The folly of fools: The logic of deceit and self-deception in human life*. New York: Basic Books.

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

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.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, 322(5898), 115–117.

Zhu, Y. L., Xi, S. S., & Wu, Y. H. (2016). Self-deception in autobiographical narrative: Individuals' self-identity strategy. *Advances in Psychological Science*, 24(12), 1917–1925. [朱艳丽, 席思思, 吴艳红. (2016). 自传体叙事中的自我欺骗: 个体的自我认同策略. 心理科学进展, 24(12), 1917–1925.]

Note: Figure translations are in progress. See original paper for figures.