Effects of Self-Deception on the Processing of Visually Presented Words

Samantha Villanueva

Colorado State University-Pueblo

A Thesis Project in Partial Fulfillment
of the Requirements for the University
Honors Program Minor

Advisors

_______________________________
Marc Pratarelli

_______________________________
John O’Connor
Abstract

Self-doubt is a form of self-deception, a process by which a person questions or doubts themselves. To assess it experimentally, the goal of this study was to examine the unconscious mechanism for self-deception. This study was the visual/verbal portion of a companion set of experiments that were based on visual stimuli. A behavioral methodology for eliciting self-deception was developed. 104 participants were shown 80 words for 2000 ms each and then given a two minute distractor task (solitaire). Afterward, they were shown a list of 40 words one at a time for 2000 ms each and were instructed to identify words that they recognized from the list of 80 study words. Unbeknownst to them, all 40 target words were randomly distributed in the list of 80 words. There was a 49.8 ms delay for responses that were incorrect, which indicated some degree of self-doubt. This delay indicates that during an incorrect response the brain is engaging (recruiting) an unconscious mechanism that allows an individual to doubt themselves. This process may represent a form of self-deception. A post-hoc experiment using a 7-point Likert scale for responses revealed that in fact all target words provided some amount of uncertainty. In those cases (trials) that resulted in an incorrect behavioral response, participants had to engage a significant degree of self-doubt because they had in fact seen all of the stimuli. Further study of self-deception using EEG and fMRI would indicate what occurs in the brain during self-deception.
Effects of Self-Deception on the Processing of Visually Presented Words

A psychological process in which a person questions or doubts what they experience is called self-doubt. Self-doubt is a form of self-deception, and is referred to as denial in the non-experimental psychology literature (Freud, 1946). The focus of the basic research in this thesis project is an experimental exploration of self-doubt in a verbal memory recognition setting to assess how individuals treat confidence in their responses.

Mele (1997) described self-deception and discussed strategies for executing the behavior. The practical example of self-deception he gives is a husband who believes his wife would never have an affair. While there is good evidence in the past of the wife not engaging in an affair, lately the wife has been coming home late and has been seen in the company of another man. Self-deception occurs when a person believes in something they want or prefer to be true. The belief does not make it true. Mele then outlines some strategies for engaging in self-deception, e.g., use of the availability heuristic, confirmation bias, a tendency to search for casual explanations, negative misinterpretation, positive misinterpretation, selective focusing/attending, and selective evidence gathering. He then gives three different types of strategies: unintentional belief, intentional activities, and intentional activities engaged with the intention of deceiving self. Mele suggests that self-deceivers must simultaneously believe in a thought and deny it by believing in its opposite.

Mele (1999) described yet another type of self-deception. In Twisted Self-Deception people convince themselves into believing thoughts or ideas that they do not believe to be true. As Pears (1985) showed, a man believes that his wife is having an affair despite his wanting to believe that she is innocent. Mele (1999) then gives three explanations for the twisted self-deception which include motivation centered, emotion centered, and a hybrid of motivation
centered and emotion centered. Motivation centered is a tendency to minimize errors based on the person’s current motivational profile, for example, a researcher concentrating on confirming instances because he is motivated to determine if a hypothesis is true or not. Emotion centered is based on insecurity and jealousy like the context of the earlier example in which a man believes his wife is having an affair despite wanting to believe that she is innocent. Clearly, his motivation to self-deceive is based on his jealously or marital insecurity. An example of the hybrid of motivation centered and emotional centered is a case in which a woman tries to explain her daughter’s death due to leukemia by believing the child caught the disease from a pet. She was motivated to believe that she was partially responsible for the child’s death and was emotionally distraught by the death of her child.

Levy (2003) discusses the predicament of an addict: they are a victim of self-deception whether or not they admit to their addiction. He characterizes self-deception with two features. The first feature is that the self-deceiver maintains contradictory beliefs concerning the subject of their self-deception. The second is that the self-deception was entered into intentionally and is maintained. He also describes two subtypes of addicts; Type 1 addicts and Type 2. A Type 1 addict denies that they have a problem with alcohol when they have good reason to suspect that they are addicted to it. Type 1 addicts have many strategies including the ability to list occasions when they did not drink, or surrounding themselves with people who drink more than they do, or thinking that their drinking behavior is normal because they have a job and a home. Type 2 addicts are what Levy refers to as an Alcoholics Anonymous addict. They recognize that they have a problem, but do not take ownership of it. Instead, they blame their addiction on biology rather than their own behavior, which attributes fault to heritability.
Tenbrunsel and Messick (2004) discussed how self-deception contributes to unethical behavior. The moral implications of a decision are able to fade due to self-deception, which is referred to as *ethical fading*. Four enablers of self-deception are discussed including language euphemisms, the slippery slope of decision making, biased perceptual causation, and the constrained representation of our self. The example given for a language euphemism is “aggressive” accounting practices versus illegal ones. Another common euphemism is calling civilian causalities “collateral damage.” The danger of using such euphemisms is that they hide moral or ethical implications of decisions that might compromise how they feel about themselves. Clearly, in the example of calling innocent people who were simply in the wrong place at the wrong time “collateral damage” allows soldiers and citizens to deny on some level that they bear any responsibility because they did not intend to kill innocent civilians.

The second enabler, the slippery slope of decision making, consists of two parts: the numbing that comes from repetition and the induction mechanism. The example given concerns changes in policy. If what was done in the past is ethical and the new practices are similar, then the new practices must be acceptable. If each step or policy change is small, then a series of changes can lead to unethical and perhaps illegal actions. The third enabler is errors in perceptual causation. Causes for errors in perceptual causation are focus on individuals rather than systems, self-interested motive in the assignment of blame, and a blurred moral responsibility involving acts of omission (Messick & Bazerman, 1996). The final enabler is the constrained representation of self (Tenbrunsel & Messick, 2004). People have no choice but to experience the world through their own bodies, rather than bodies of others. Because of this even when people try to put themselves in the place of another, the best they can do is to try to imagine what another would experience from their own perspective.
Browndyke et al., (2008) used fMRI to compare data from normal recognition and purposeful malingering in participants. Listed were many studies that used an increase in reaction time to determine whether there was a process occurring, but they do not address the neural substrates behind deceptive behavior. The goal of their study was to identify the neural architecture for deception. There were only seven participants in the study, which is not entirely uncommon in fMRI methodology. Participants were presented with black-and-white two dimensional drawings against a white background. The drawings of the namable objects were taken from two separate tests: one was the Test of Memory and Malingering and the second was the Snodgrass and Vanderwart visual object stimuli set. There were two encoding tasks run each with 50 items. Each item was shown for two seconds followed by a static visual crosshair inter-stimulus interval of varying duration. In the first set the participants were asked to learn and remember all object pictures for a later recognition test. Participants used buttons to indicate whether the stimulus shown was new or old with regard to the 50-item study set. In the second run (malingering) participants were asked to respond in a different manner than before. In this instance they were to imagine they were slightly injured in a car accident and were feigning a memory impairment for financial gain. Participants were told not to be too obvious as there was a risk of punishment. The behavioral results from Browndyke et al., showed that recognition discrimination for the malingering condition was markedly impaired. The results for the normal condition were superior. In the normal condition target hits and malingered target misses had the most significant differences, which appeared in the right Inferior Parietal Lobule and left dorsomedial prefrontal cortex. Smaller significant clusters were also found in the right lingual and left precuneus. Correctly identifying new stimuli in the normal condition was associated with large clusters in the left lingual and cuneus regions. In the malingering condition false
alarms for new stimuli was positively associated with activity in the right middle temporal gyrus. The middle temporal gyrus was noted as being significantly more active during deceptive endorsement than the normal condition when responding to new stimuli. Their results provide compelling evidence for what occurs physiologically when a person attempts to deceive another individually. It leaves no doubt that the intention to deceive leaves a cortical signature that may help understand the timing of such mental operations in future research.

Visu-Petra et al., (2012) studied individual differences in executive functioning and time based detection of concealed information. Deception requires the use of higher-order cognitive thinking control mechanisms. These mechanisms are required because deception is a complex task involving the choice to be deceitful rather than truthful, to overcome the awareness of the truth, to suppress memory and expression, to generate alternatives that are plausible, to maintain consistency of the lie, monitor personal behavior, and the audience’s reaction (Gombos, 2006). Participants were assessed in the areas of cognitive abilities and personality dimensions related to deceptive behavior using multidimensional executive functions and anxiety measures (Visu-Petra et al., 2012). The assessments were done using the Cognitrom Assessment System. A mock crime scenario was used. Participants were assigned in one of two groups, guilty or innocent. The guilty group was aware of six critical details and the innocent group was not aware of them. Guilty participants were required to steal a CD containing exam questions for an upcoming exam. Guilty participants were warned to be cautious but to not be seen behaving in a suspicious manner. They were asked to go to an office, locate a laptop bag, find a CD case within it, open the CD case to verify it was the one with the exam questions, and note the 100 RON banknote inside (RON is the currency used in Romania). The items were taken from the guilty participants and then all participants received instructions stating that they were suspects
in a theft and would undergo a behavioral test and a polygraph test to assess their involvement in the mock crime. All participants were re-administered an anxiety test to determine if there was an increase. After the polygraph test participants were given a reaction based concealed information test. Three groups of study pictures were used; probes (items from the mock crime), targets (to be detected items) and irrelevants (four for each probe). Participants were instructed to respond Yes to targets and No to probes and irrelevants. Guilty subjects had poorer accuracy on the probes compared to irrelevants, yet for innocent participants there was no significant difference. Reaction times for probes were slower than irrelevants in the guilty participants. No significant interrelations were found between executive functions and outcomes from the recognition memory test in the innocent group. In the guilty group spatial working memory was positively related to reaction times for irrelevants, probes and targets. Subjects with better spatial working memory took longer to correctly classify all types of items and were better at recognizing targets. Time of anxiety assessment and guilt condition were not significant, meaning it was found to be the same in both innocent and guilty participants.

Audi (1982) gives the example of a patient who is dying of cancer deceiving herself. She avoids letting her doctor give her a prognosis, discusses her plans for the long future and her recovery. This example is similar to examples given by Kubler-Ross (1969) in which a patient believes that her files were mixed up with another patient’s and went shopping around for doctors who would give her a better explanation of her symptoms. She repeatedly asked for re-examinations because she believed that an error had been made. Kubler-Ross also gave an example of another patient, a young mother who denied her illness for as long as possible even when hospitalized. She claimed that a faith healer had cured her illness. Although she believed that she was not ill, she never tried to leave the hospital. Audi states that weakness is a
contributor to self-deception (1982). There is an action, intention, want or perhaps belief formed or maintained against one’s better judgment. There could also be failure to perform an action, intention, want or belief when one’s judgment requires it. He claims that people become self-deceived partly or largely through actions, and as a result of being self-deceived, which is a circular argument.

Audi also discussed moral responsibility and self-deception. There were three issues: moral responsibility for getting into self-deception, for remaining in it, and for acting out of self-deception. With regard to getting into self-deception one must do more to suppress a belief than to acquire one unconsciously. A person can act intentionally to get into self-deception or avoid it. Someone who is in self-deception mode rarely acts intentionally to remain in it or to get out of it.

There is physiological literature on what occurs when one attempts to deceive another, but not one’s self. Behaviorally in the present thesis project, there should be delays in response time for denied (self-deception) responses as compared to correct trials. There should be fewer denied responses as compared to correct verbal responses and nonverbal symbol responses (companion study), which will be discussed later. To examine certainty, the Likert Rating experiment should produce a similar pattern to the response time version. In the present study the broad goal was to determine if there was any experimental behavioral brain-based cognitive mechanism for self-deception.

Method

Participants

104 undergraduate students attending a small Southwestern university ranging in age from 17 to 53 (M = approximately 23.5 years), were selected using a convenience sample.
females and 40 males of varying ethnicities participated in this study. All conditions were repeated measures, and all participants took part in the study individually. That is, there was no group administration.

**Materials**

A list of 80 study words (Table I) was generated using the “Frequency Analysis of English Usage” (Frances & Kucera, 1982). From the list of 80 words a second list of 40 target words (Table 2) was generated from the list of 80 study words. Words were four, five, six or seven letters long. The study and target lists were balanced for word length, word frequency of use, and grammatical part of speech. In the study list there were 45 nouns, 13 adjectives, 2 prepositions, 17 verbs, 2 adverbs, and 1 post determiner. In the target list there were 21 nouns, 8 adjectives, 1 proposition, 8 verbs, 1 adverb, and 1 post determiner. The average frequency of use was 120.59 for the study list and 128.78 for the target list. When organizing the words as they would appear to participants care was taken to ensure that words like traffic and women were not near each other because together they have a different meaning than they do in isolation. Care was also taken to ensure that words that were phonetically, graphemic or semantically similar were kept as far away from each other as possible.

**Procedures**

Superlab Pro software was used to present the stimuli on a conventional desktop computer. The first set of tests were set up with the 80 study words presented first, followed by the 40 target words. The second test had the 40 target words presented first, followed by the study list of 80 words. The first test phase had participants look at the study list first; each word was shown for 2000 ms, and then solitaire was used as a distractor task for two minutes to prevent continued rehearsal. Participants were then given the list of 40 target words. Each target
was shown for 2000 ms and participants were asked to respond whether or not they had seen the word in the previous study list. In reality, and unbeknownst to the participants, all of the words in the target list had appeared in the study set. The presentation was assembled in this way to make students think that it was unlikely that all of the target words were seen in the study list. This would be similar to taking a multiple choice test and having all of the answers be “c.” This is very unlikely and students would doubt that it would ever occur. The second test phase was conducted in the same manner but students were given the target list first and the study list second. This was done to assess any confounds with respect to the individual stimuli. They responded whether or not they had seen the study words before and only half of them were seen in the target list. The third test was set up similarly to the first test; participants were presented with the study list of 80 words, followed by the distractor task, and then were shown the list of 40 target words. In this test participants were asked to respond using a seven-point Likert scale. A score of one meant absolutely sure that the word was not in the study set, while a score of seven meant absolutely sure that the word was in the study set.

**Results**

**Reaction times**

An ANOVA was used to compare correct and denied trials. The average reaction time for correct trials was 859.76 ms (SD = 124.00 and SEM = 2.82) and was 909.53 ms (SD = 176.28 and SEM = 4.01) for denied trials (Figure 1). There was a marginal effect of reaction time between correct and incorrect (denied) trials: \( F(1.86) = 2.35, p < .129 \). A power analysis suggested that twice the data set would be necessary to reach significance at \( \alpha = .05 \).
Accuracy

Percent accuracy was compared using a depended samples $t$-test. The percent accuracy for correct trials was 47.82% and 25.52% for incorrect trials. This means that 47.82% of the time participants had a correct answer, and 25.52% of the time participants had an incorrect (denied) trial. The correct and incorrect trials do not add up to 100% due to nonresponses. There was a significant difference between percent accuracy for correct and denied trials, $t (1,43) = 9.184, p < .0005$ (Figure 2).

Accuracy Results Using Likert Ratings of Confidence

For correct trials the average Likert rating (seven-point scale) was 5.27, the average for denied trials was 4.32 (Figure 3). There was a significant difference between Likert ratings for correct vs. denied trials, $F (1, 37) = 77.45, p < .0005$. There were only 39 words in the target list, due to an issue with the program.

Discussion

It is virtually axiomatic in the speech sciences that “speech is special.” It means verbal processing has a special place in the human brain because it is the primary form of symbolic communication. In contrast nonverbal symbols are flexible and generic. In the companion study the symbols were meaningless. In contrast to the nonverbal symbols, the words in the present study had a strong semantic context. Semantics gives words an added advantage to combat ambiguity and self-deception. Therefore, it should be expected that the present results show words are less ambiguous to participants, and thus more difficulty to be mistaken and denied. As an aside, the more unique and distinctive a word is perceived to be the easier it is to recognize as having been studied (Hunt & McDaniel, 1993). Therefore, distinctive words should be more difficult to confuse and cause self-doubt or self-deception.
The delay in reaction time for denied trials indicates that there is a mechanism in the brain for self-deception. If there was no mechanism there would not be a delay in reaction time because nothing different would be occurring in the brain between correct and denied trials. This mechanism is activity in the brain that is different from when there is a correct trial. The denied trials are self-deception because the participants were given two choices but only one was correct; all stimuli from the target list were present in the study list. There was only one correct answer but two choices created self-doubt. Why would an option be given in a test if it was not an answer for one of the questions? Participants thought since there were two answers both of them would be used. It is similar to a multiple choice exam having only one answer thought the exam, for example “C”. It would be hard to believe that all of the answers are the same, and so many taking the test would change some of their answers.

There were some differences in reaction times in the current study and the companion study. In the companion study the difference in mean reaction time between correct and incorrect responses was about one-fifth of a second (191.3 ms). In the current study the difference in reaction time between correct and incorrect responses was a mere one-twentieth of a second (49.8 ms). The difference in reaction times between the two studies is due to semantics, the words have semantics while the symbols lack semantics because they are generic.

The companion study was a nonverbal symbol experiment set up in the same manner as the current study. Participants in the nonverbal symbol experiment were shown 40 symbols and then given a two minute distractor task (newspaper). Afterward they were shown a list of 40 symbols and were instructed to identify symbols they recognized from the study set. Both the current study and the companion study showed a delay in reaction time in denied responses. The delays both indicate that perhaps a common mechanism for self-deception exists to cause a
common delay in processing. Comparing the results from the current study and the companion study using ANOVA found a main effect for both accuracy (correct vs. denied) using verbal vs. nonverbal symbols as the independent variable and reaction times as the dependent variable, $F(1,184) = 33.11, p < .005$ (Figure 4).

More importantly, there was an interaction effect between accuracy and verbal vs. nonverbal symbols, $F(1,184) = 8.52, p < .004$ (Figure 5). Responses to words was much faster because there is less ambiguity due to the benefits of semantics. The average response time for words was 884.6 ms, while the average response time for symbols was 1024 ms. Once again, the difference in mean response times is due to semantics, as the symbols were generic. Overall in both studies there was a delay in response time for denied trials.

Browndyke et al., (2008) used fMRI to compare data from two groups of participants normal and malingering. Participants in the malingering (deceiving group) had different brain activity than normal participants. If more data collection were to be done with the present study using fMRI, it could be possible to see a difference in brain activity when participants are shown words from the target list vs. words not in the target list. Repeating this study with fMRI could be used to show the differences in brain activity when self-deception occurs and possibly where the mechanism for self-deception resides.

Visu-Petra et al. (2012) also used reaction time. They found that reaction times were slower for probes (items from the mock crime) than irrelevants in guilty participants (those who were asked to deny involvement in a mock crime). When participants were actively engaged in deceiving another person there was a delay in reaction time. The current study had a similar outcome, denied responses (self-deceived) had longer reaction time than correct responses. When participants in the current study actively deceived themselves there was a delay in reaction
time. Participants in both studies were shown the information on a computer using Superlab software and were asked to respond using corresponding keys with the index fingers. In Visu-Petra et al., (2012) the image remained on the screen until the participant made a response. In the current study participants were given 2000 ms to make a response. After 2000 ms no response was recorded and the next stimulus was shown regardless if the participant gave an answer.

The limitation of the current study is how difficult it is to set up a scenario that causes self-deception. There is also a lack of empirical research on self-deception. There have been studies that focus on what occurs when a person tries to deceive another, but no others than the current study on self-deception and its companion, perhaps due to the difficulties of setting up a situation in which self-deception can be measured experimentally. The current study uses reaction time to measure self-deception, but it does not show where or what is occurring. Follow up studies using EEG and fMRI would need to be done in order to know what occurs during the delay in reaction time.
References


Table 1

*Study List*

<table>
<thead>
<tr>
<th>Study List</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
</tr>
<tr>
<td>Coal</td>
</tr>
<tr>
<td>House</td>
</tr>
<tr>
<td>Short</td>
</tr>
<tr>
<td>Guide</td>
</tr>
<tr>
<td>Mice</td>
</tr>
<tr>
<td>Purple</td>
</tr>
<tr>
<td>Apple</td>
</tr>
<tr>
<td>Picture</td>
</tr>
</tbody>
</table>

**Part of Speech**

<table>
<thead>
<tr>
<th>Nouns 45</th>
<th>Adjectives 13</th>
<th>Prepositions 2</th>
<th>Verbs 17</th>
<th>Adverbs 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Determiner 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Average Frequency 120.59**

*Note.* The study list was shown first to participants.
Table 2

**Target List**

<table>
<thead>
<tr>
<th>Words</th>
<th>Goal</th>
<th>Lawn</th>
<th>Found</th>
<th>Suppose</th>
<th>Farther</th>
<th>Latter</th>
<th>Choice</th>
<th>Front</th>
<th>Weather</th>
<th>Shown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dice</td>
<td>Series</td>
<td>Relief</td>
<td>Project</td>
<td>Wrong</td>
<td>House</td>
<td>Ocean</td>
<td>Driver</td>
<td>Recent</td>
<td>Explain</td>
</tr>
<tr>
<td></td>
<td>Purple</td>
<td>Yellow</td>
<td>Kettle</td>
<td>Better</td>
<td>Group</td>
<td>Coffee</td>
<td>Argue</td>
<td>Issues</td>
<td>Threat</td>
<td>Casual</td>
</tr>
<tr>
<td></td>
<td>Juice</td>
<td>Inside</td>
<td>Minute</td>
<td>Assess</td>
<td>Risky</td>
<td>Traffic</td>
<td>Local</td>
<td>Board</td>
<td>Work</td>
<td>Bank</td>
</tr>
</tbody>
</table>

**Part of Speech**

<table>
<thead>
<tr>
<th>Nouns</th>
<th>21</th>
<th>Adjectives</th>
<th>8</th>
<th>Preposition</th>
<th>1</th>
<th>Verbs</th>
<th>8</th>
<th>Adverb</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Determiner</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average Frequency 128.78

*Note.* The target list was shown to participants second, after the study list and distractor task. It was balanced to ensure that the average word frequency was similar to that of the study list as well as the part of speech.
Figure 1. Mean reaction time in ms for correct vs incorrect responses. There was a marginal effect. $F(1, 86) = 2.35, p < .129$
Figure 2. Accuracy for correct and incorrect responses, the difference between correct and incorrect responses is statistically significant.

\[ t (1, 43) = 9.184, p < .0005 \]
Figure 3. Results from the seven-point likert scale ratings for confidence, correct trials received a higher rating on seven-point likert scale compared to incorrect trials. There was a significant difference between the two groups. $F(1, 37) = 77.45, p < .0005$
Figure 4. Comparison of overall reaction time between nonverbal symbols study and visual/verbal study. There was a statically significant difference between the studies, due to semantics. $F (1,184) = 33.11, p < 0.005$
Figure 5. The interaction effect between accuracy and experimental modality between nonverbal symbols study and visual/verbal study was statistically significant. The difference in reaction time is due to semantics. $F(1,184) = 8.52, p < 0.004$