SELF-DECEPTION WITHOUT PARADOX

Dante A. Cosentino

November, 1980
ABSTRACT

SELF-DECEPTION WITHOUT PARADOX

In this paper a view of self-deception is given which eliminates the paradox usually associated with self-deception (i.e., the self deceived believes both p and not-p at the same time).

Self-deception is distinguished from ignorance, false belief, wishful thinking, and reluctance to believe. Through an analysis of ordinary language, the role of knowing and believing in self-deception is examined as well as the notion of the self-deceived person "persuading himself to believe." The role of intention and the function of "self" in "self-deception" is analyzed through a discussion of evidence and interpretation.

It is shown that the self-deceived person does not know or even believe p, but is, nevertheless, properly characterized as "self-deceived." It is argued that the self-deceived person believes not-p and does not know, or even believe, p. The paradox turns out to be more apparent than real.
Self-Deception Without Paradox

Although the phenomenon is familiar, philosophical attempts to explicate the concept of self-deception have run into difficulty because of an apparent paradox, i.e., a self-deceived person is both deceiver and deceived. But how can he believe both \( p \) (as deceiver) and believe not-\( p \) (as deceived) at the same time? This apparent paradox is, what one writer on the subject calls, "the problem of self-deception."\(^1\) There have been various approaches to this problem and at least three can be readily distinguished: the "two levels of awareness" approach,\(^2\) the "in a sense he believes, in a sense he does not" approach,\(^3\) and the "volition-action" approach.\(^4\) All three approaches fundamentally assume that there is a paradox to be solved, i.e., the self-deceived believes both \( p \) and not-\( p \). I will argue that there is no such paradox. The intent of this paper is not to solve a paradox but to remove the ground on which the paradox rests.


\(^2\)Ibid.

\(^3\)T. Penelhum, "Pleasure and Falsity," in Philosophy of Mind, ed., Stuart Hampshire, Harper & Row, New York, 1966, pp. 242-266. The name I give to this approach is not Penelhum's, but I am using this name for the type of approach which rejects the other-deception model and talks in terms of a qualified knowing or believing, e.g., in Penelum's case, he says, Someone in this state does partially satisfy the criteria for belief and also those for disbelief." p. 258.


\(^5\)Demos: "Self-deception exists...when a person lies to himself, that is to say, persuades himself to believe what he knows is not so. In short, self-deception entails the B believes both p and not-p at the same time." Demos, p. 588.
I am proposing a thesis which claims that a person whom we characterize as self-deceived believes not-p and does not know, or even believe, p. Furthermore, I claim that the self-deceived does not engage in an act of deception, e.g., he does not persuade himself to believe not-p, but, nevertheless, we may properly call him "self-deceived."

This interpretation will rid us of the apparent paradox without losing the phenomenon. The following sections will give an explanation of this view and attempt to answer the following questions which might be raised:

1. If the self-deceived does not know p, why do we say "he deceives himself"? Is this not just a case of ignorance or false belief?

2. How is self-deception distinguished not only from ignorance and false belief but also from (a) other-deception, (b) wishful thinking, and (c) reluctance to believe?

3. If the self-deceived is not deceiving himself, then what is the function of "self" in "self-deception"?

4. If there is no agent who engages in an act of deception, why even call it a case of deception at all?

5. Does this interpretation correspond to our everyday talk about self-deception?

While I agree with other writers on self-deception that in "A is self-deceived" it is true that A is deceived, I do not agree that he deceives himself, if part of the meaning of "he deceives himself" is taken as "he knows p."

---

Penelhum: "The concepts of self-deception has seemed to some to generate paradoxes, to make us claim, for example, that a man both does and does not believe the same proposition.... The only way of avoiding the reinstatement of the paradox here is to accept the fact which paradoxical renderings feed upon: that self-deception is a conflict state." Penelhum, p. 258.

Fingarette: "It is correct to say that, even if we assume no motive, we would call a person self-deceived if he persuaded himself to believe what in his heart he knows is not so." Fingarette, p. 28.
An immediate, obvious objection is this: If "self-deceived" does not mean "he deceives himself," which seems to entail that he knows p (and is what we usually take it to mean), what else can it mean? It is not enough to say it just means "he is deceived" because we want to know who is deceiving him and how he is being deceived. If he is deceived and no other is deceiving him, then is he not deceiving himself? In other words, if one wants to maintain that the self-deceived does not know p and yet wants to retain the expression "self-deception," it will have to be shown that A does not know p and that we can still make sense of the expression "A is self-deceived" (and/or "he's deceiving himself").

A partial answer to these objections is that while it is true that when we say "A is self-deceived" at least part of what we usually mean is that no other is deceiving him, it does not follow from that that he is deceiving himself in the sense that he knows (and believes) p (but "persuades" himself of not-p). What the expression "self-deceived" does point out is that no other person is trying to deceive him, but A is, nevertheless, deceived: he believes not-p. But to conclude from this that A must know p and that there is not only one who is deceived but an agent (who commits an act of deception) as well is unwarranted. It stems from the uncritical acceptance of "he deceives himself" as somehow analogous to "A deceives B" (where A and B are two different persons), so we are misled into analyzing it in the same way. In "A deceives B" there is a deceiver who knows ("A"), an act of deception ("deceives"), and a person who does not know and upon whom the act is perpetrated ("B"). From the similarity of form it looks as though "he deceives himself" might be susceptible to the same kind of analysis, i.e., a deceiver who knows ("he"), an act of deception ("deceives") being perpetrated on the deceived ("himself"). Thus from the joining of "A deceives B" to "He deceives himself" emerges the paradox. My analysis of "he deceives himself" will become obvious as the paper proceeds, but to anticipate, I will say that when we say "he deceives himself" it is not the opposition between what he knows and what he believes, but, rather, it is the opposition between what we know and what he believes which we find so bothersome.6

6 What we know is something the self-deceived ought to know and would know were it not for certain emotional needs of his, i.e., his belief is engendered and sustained by certain emotional needs (e.g., his desires, anxieties, expectations, etc.). I would agree with Patrick Gardiner's comment on this aspect of self-deception when he says:
Even writers who claim to have discarded the other-deception model reintroduce it when they take "he deceives himself" to mean that he knows (at least in some sense) p and believes not-p, and in their use of expressions like "he persuades himself" to believe (or "he makes himself" believe) the contrary of what he knows.  

There are two distinct but related mistakes involved in the apparent paradox. The first mistake is to say that the self-deceived both believes p and believes not-p. The second mistake is to assume that where one is deceived there must be an agent who engages in an act of deception. This second mistake is related to the first in that both use other-deception as their model. In other-deception there is a victim (the one deceived) and a villain (the one who deceives). The victim does not know p and believes not-p. The villain does know and believe p. Hence the talk, when applied to self-deception, of the self-deceived as one who believes p and believes not-p at the same time. In other-deception it is also clear that there is an act of deception, so now we have victim, villain, and act of deception, which is the model used for self-deception and becomes the source of the second mistake. It is this other-deception model which gives us the picture of believing both p and not-p at the same time. Once we break the hold of the other-deception model we will see that it is not the case that the self-deceived both knows (and believes) p and believes not-p. Furthermore, it is not the case that where one is deceived there must be a deceiver who engages in an act of deception.

"It is, for example, not clear what could be meant by, or what justification there could be for, speaking of somebody as deceiving himself if it were at the same time contended that what he was said to be deceiving himself about was a matter of total indifference to him, in no way related to his wants, fears, hopes and so forth: could we, e.g., intelligibly talk about 'disinterested' or 'gratuitous' self-deception."


J. Canfield & D. Gustavson, for example, who explicitly reject other-deception as a model, say, "...it is possible to interpret self-deception as a special case of self-command, i.e., as making oneself believe something or forget something." See "Self-Deception" in Analysis, Vol 23, pp 32-36. This quote is on page 33 (their italics).
It is assumed, rightly I think, that in genuine cases of deception (i.e., where one intends to deceive another, not necessarily in cases where one unintentionally deceive another) the knowledge of p must be prior to a successful act of deception about p. It is important to note that where one intentionally deceives the knowledge of p must be active at the time of deception, e.g., it cannot be something which the deceiver once knew but has forgotten at the time of the act. So the deceiver knows p before and during the act of deception. It is also assumed (and again I agree) that the self-deceived has the same evidence for p which we all have, and in this respect it is unlike other-deception. But if A, who is self-deceived, has the same evidence as the rest of us (from which we draw the conclusion p) and does know p, then it seems that he has already seen the evidence and drawn the correct conclusion but has now given the evidence a second reading, misinterpreted it this time, and concluded not-p.

However, the fact that he misinterprets (and/or ignores) evidence is not enough to show that he is self-deceived. Most of us, at some time, misinterpret facts and draw mistaken conclusions. Scientists, historians, and detectives, at times, misinterpret and/or ignore evidence but we do not, on that basis alone, call them self-deceived. Those holding the view that the self-deceived knows p would have to show that he knows p while in a state of self-deception, and this is what is usually assumed but not shown. In defense they might say, "Well, you agree that he sees the evidence, so he must know p." But the evidence is evidence for the proposition "p" (what is the case) not for the proposition "he knows p" (which is what the claim is). Nor will it do to show that the self-deceived knew p at some time in the past. It may be obvious, but it should nevertheless be kept in mind, that we must be aware of tense-shift in discussions of self-deception. Whether the self-deceived knew p in the past is irrelevant, what is relevant is whether he knows p now, i.e., at the time he is in a state of self-deception. This does not exclude past (or future) knowledge of p. Normally, when we talk about self-deception, it is

---

8Henceforth when I use "misinterpret the evidence," I mean to include in that expression cases where the person might misinterpret and/or ignore at least some of the evidence.
what is presently going on that we are concerned with, not what happened in the past or may happen in the future. To talk about the fact, if it is a fact, that A did know p at some time in the past, as proof that (despite his present denial and behavior) he does know p, is to make an illegitimate move. To keep the paradox of self-deception alive one must show that the self-deceived person concurrently knows p and believes not-p.

The oblique move around "he knows p," as in the "in a sense he knows" approach, runs into other difficulties; what exactly is the sense in which he knows p? He does not say he knows p. Does he then, in his non-verbal behavior, show that he knows p? This may not be as clear-cut, but, in general, the non-verbal behavior of the self-deceived seems to fit his verbal behavior. If it did not (e.g., if A said that his son was not a thief but constantly locked up all his family valuables only when his son was in the house, or if B said his wife is faithful but hires a private detective to watch her), we would probably say that he is being hypocritical or, more charitably, that he is pretending not to know. The genuine case of self-deception, the one which puzzles us, is not where his actions "say" something different than his words but where his actions fit his words, where what is apparent (p) to every reasonable person (with the same evidence) seems not to be apparent to the self-deceived person.

There might be an objection raised here, i.e., is it not possible that the actions of the self-deceived go against his claim to believe not-p? This possibility will be discussed at the end of this section.

Let us first look at what might be a common case of self-deception: Mrs. Smith who believes her son, Timmy, is a good boy (not-p) when he is really a scoundrel (p). The evidence shows what we know to be true, e.g., Timmy is a bully, he lies, steals, and cheats. We see Timmy hitting smaller children without provocation. We hear him lie that they hit him first. We watch him steal toys from other children and then say that the children gave him the toys. We watch him cheat while he is playing games. Mrs. Smith sees the same evidence we do but steadfastly maintains that Timmy is a good boy. She certainly seems to be sincere when she tells us that Timmy is a good boy, and we know that she is, in other circumstances, a very reasonable and honest person. We are puzzled. We wonder why she can't see that Timmy is a scoundrel, since it is obvious (to us). We might say something like, "she sees the same things we do, she must know it, all the evidence is there." We finally decide
that "she is deceiving herself" and that "she must know" and
do not inquire into how she is deceiving herself or what it
means to say "she must know."

Mrs. Smith is self-deceived in that she believes not-p
even though p is true, she interprets the evidence in such a
way as to conclude not-p, and not-p is a desirable belief.9
How did she come to believe not-p?

It is possible that initially Mrs. Smith believed not-p
and had some evidence for it. Timmy was a happy baby and
developed normally as a pre-school child, in such a way that
one could characterize him (accurately) as "a good boy."
When he started school, however, he got into bad company,
and learned to lie, steal, cheat, and bully. At this point
"Timmy is a good boy" becomes false and "Timmy is a scoundrel"
becomes true.10 Mrs. Smith continues to believe Timmy is a
good boy (not-p). What she now sees (i.e., the facts which
the rest of us accept as evidence that Timmy is a scoundrel)
she interprets incorrectly because she believes not-p and not
because though she knows p she has somehow succeeded in "mak­
ing herself believe" not-p. We are not satisfied with her
constant "explanations" of Timmy's behavior and feel that
"she must know" p. To us she seems only to be trying to

9 Writers on self-deception seem to agree that the self-
deceived consistently interprets the facts differently from
the rest of us. But if someone consistently misinterprets
the evidence, does this not show that he does not know (is
it not something like the "seeing as" case, i.e., someone
consistently sees x as y. Would we, after he sees x as y
for the hundredth time, and never sees it as x, still want to
say that he must see x? If so, on what grounds?)

10 In the beginning her belief is true, circumstances change
but her belief does not. In some case of self-deception the
belief may be false to begin with but in those cases it is
probably because there is no evidence available to the person
which goes against the belief, e.g., a person may believe his
(her) spouse is faithful but in fact the spouse is unfaithful.
The spouse may even have been unfaithful since the day they
were married, but the self-deceived has no evidence available
to indicate that this is the case. But even in cases such
as this, where the belief was never true, there is usually
some reason for the false belief (possibly marriage vows in
this case).
"persuade herself" of not-p, and we characterize her as "self-deceived." It has not been shown, however, that Mrs. Smith actually knows p (although, admittedly, she is in a position to know p), or even that she is attempting to persuade herself of not-p. What is clear is only that she believes not-p. Mrs. Smith is attempting to persuade the rest of us of what she "knows," her explanations are not an attempt to persuade herself of not-p, for she really believes that. She is trying to persuade others of what she thinks she knows. We do not usually have to persuade ourselves of what we already (think we) know.

In Mrs. Smith's case her other verbal behavior (i.e., other than her claimed belief in not-p) and her non-verbal behavior correspond to her belief in not-p (her non-verbal behavior: she rewards Timmy "for being good," she never spanks him, etc.). But what about a case of self-deception where the self-deceived person's non-verbal behavior seems to go against his claimed belief? The case of the alcoholic seems to be such a case. He is neither a hypocrite nor is he pretending when he says, "I am a social drinker" (not-p), but, nevertheless, his actions do seem to "say" something different than his claim, i.e., his actions show that he is an alcoholic (p). His claim, which seems sincere, that there is very little difference, if any, between his drinking and that of others is contradicted by his behavior. It looks as though his behavior is not in accord with his belief in not-p.

I agree that there are cases of self-deception (such as the case of the alcoholic) where the behavior of the self-deceived is not in accord with his stated belief. However,

---

11 We may even say, "she does not want to believe that her son is a scoundrel," or she does not want to see that her son is doing those things we mentioned," and these expressions can be read as indicating that our ordinary language suggests that she does know p (but does not want to believe it, or is trying somehow to persuade herself of not-p). I think there is an alternative way of reading those expressions which indicate not that she knows p but that she has a vested emotional interest in not-p, i.e., she wants to believe not-p and does believe not-p. The fact that someone wants to believe not-p and even says that he does believe not-p are rather unusual grounds for claiming that he knows p. Another related expression which seems to support my contention that she does not know p is the one in ordinary language which says, "it's in front of her eyes but she does not see it," i.e., the evidence is there but she does not take it as evidence for p. All of these expressions, rather than supporting the claim that the self-deceived knows p, seem to lean the other way, i.e., she wants to, and does, believe not-p.
it is important to note here that the evidence for \( p \) is his non-verbal behavior whereas in the previous example we were able to separate the evidence for \( p \) (e.g., Timmy's behavior) from the verbal and non-verbal behavior of the self-deceived person (Mrs. Smith's stated belief in not-\( p \) and her acting towards Timmy and others as if she believed that Timmy was not a scoundrel).

The alcoholic's non-verbal behavior does not show that he believes something other than what he claims to believe. What his non-verbal behavior shows is that he is an alcoholic, not that he knows he is an alcoholic. When the evidence for \( p \) (viz., his behavior) is pointed out to him, he misinterprets, he rationalizes, in the same manner as Mrs. Smith does. For example, if we point out that whenever he receives his paycheck he always pays his bar and liquor store bills before any other bills he might respond with, "the bartender is a good friend and trusts me, I don't want to let him down," or perhaps, "the liquor store will cut off my credit if I don't pay every month, but I can always skip a credit card payment or the rent payment without getting my credit cut off or being forced to move." He does not say, "I'm an alcoholic, so I take care of the bar bill and liquor bill first."

Neither his verbal behavior nor his non-verbal behavior shows that he believes \( p \) (although his non-verbal behavior is, for others, evidence for \( p \)).\(^{12}\) So the case of the self-deceived person whose actions seem to go against his claimed belief is subject to the same analysis as the previous example once we understand that in the case of the alcoholic his non-verbal behavior is the evidence for \( p \). Once the alcoholic acknowledges what the rest of us know, viz., that he is an alcoholic, then he is no longer self-deceived.\(^{13}\) We should

\(^{12}\) Of course, alcoholism is a complex phenomenon which is, of necessity, being over-simplified here. However, that an alcoholic does act this way is obvious to anyone who has known an alcoholic or even given a cursory look at the literature on alcoholism. For example, Marty Mann, speaking about the excuses made by the alcoholic for his drinking, speaks of, "the symptom called 'rationalization' or 'the alibi structure'." See Marty Mann's New Primer on Alcoholism, Holt, Rinehart and Winston, New York, 1958, page 21.

\(^{13}\) When the alcoholic knows that he is an alcoholic his self-deception ends, although he is still an alcoholic. Most experts in the field agree that alcoholism is a disease which cannot be cured but only arrested. I am using as my model here E. M. Jellinek's "gamma alcoholic." See E. M. Jellinek, The Disease Concept of Alcoholism, Hillside Press, N. Haven, 1960.
also note that one of the reasons that we are inclined to say that the alcoholic is self-deceived (rather than, for example, calling him a hypocrite) is because we assume that he does not realize that he is an alcoholic, i.e., he does not know p.

II

If the self-deceived does not know p how does he differ from one who is just ignorant or one who merely holds a false belief?

Self-deception is like ignorance in that the self-deceived does not know p and, like ignorance, it may be not only that he does not know p but that he may have known p at one time but has now forgotten it, in which case he does not know p, he is ignorant of p. Self-deception sometimes has the appearance of ignorance, but in various ways the self-deceived person is different from one who is simply ignorant. When we say of someone that he is ignorant about a particular subject we usually mean he is uninformed or misinformed about the facts. The self-deceived, however, is neither uniformed nor misinformed about the facts relating to p. If a reasonable person is ignorant with respect to p, that ignorance can usually be rectified by his being apprised of the facts. This is not the case with the self-deceived.

Another distinction between the one who is ignorant of p and the self-deceived is that the latter has a desire or need for not-p to be true. The former normally does not have such a need or desire. When the evidence is presented, the one who was ignorant accepts p whereas the self-deceived person gives us counter-arguments or rationalizations. The ignorant one does not persist in his denial of p, the self-deceived does.

14

There is a colloquial usage where the word means a person is just unable to understand, i.e., he is mentally incapable of understanding (e.g., "he is just ignorant, he will never understand no matter how much you explain it to him"), but the self-deceived is not ignorant in this sense. We are amazed by his reactions to the evidence because we think that he is normally a reasonable person and is capable of understanding once his error is pointed out to him.
The distinction between false belief and self-deception follows lines similar to that between ignorance and self-deception. Normally when the evidence which is contrary to our false belief is presented to us, we do not point to what was previously the case, or misinterpret the evidence, in order to hold on to our false belief (although we may point to it to show how we came to believe falsely). A reasonable person, when faced with evidence which all others around him accept, but which is contrary to his belief, does not consistently misinterpret the evidence and insist that he is right but simply drops the belief. When I say that the self-deceived is acting unreasonably by not dropping the belief, I do not mean that he knows p but will not accept it, but that a reasonable person in that situation would know p, i.e., the self-deceived should know p, but does not.\footnote{A word about mistake in self-deception. If there is a mistake involved it is not a simple mistaking one thing for another but, as Frederick Siegler has pointed out, it is a mistake at the level of inference. Siegler however, takes the mistaken inference as leading to the false belief (and he espouses the "he persuades himself" argument) which I think is wrong. He says, "But in self-deception this distortion is in the manipulation or twisting of evidence that we get, and it is at the level of inference which leads to beliefs." See F. Siegler, "An Analysis of Self-Deception," Nous, Vol II, No 2, May 1968. The mistaken inference, as I see it, is an attempt to justify the false belief, which the self-deceived already has. So rather than leading to the false belief, the mistaken inference is a consequence of that belief.}

It might be argued that on my interpretation we have obliterated the distinction between wishful thinking and self-deception, i.e., in wishful thinking one also holds a false belief because of some emotional need that the belief is true. It is this emotional need which gives the wishful thinker his motive: if this need did not exist he would not hold this false belief and in these aspects it looks as if what I have described as self-deception could also be wishful thinking. Although there are these similarities, there are also differences. One of the differences between wishful thinking and self-deception is that the wishful thinker when faced with evidence which is contrary to his belief, will accept that evidence and its natural implications. As we have seen, the self-deceived person does not accept the evidence. In fact, he reaffirms his belief in not-p, and tries to justify it by misinterpreting the evidence.
Another difference is that the self-deceived person must hold a false belief whereas in wishful thinking, at least in some cases, there is no false belief involved. Consider the following: Someone expresses a desire to win the Irish Sweepstakes and even talks about what he will do with the money. We say to him, "It is foolish to talk about what you will do with the money which you have not won, and the odds of your winning are probably a million to one." He replies, "I know that, it is just wishful thinking on my part." This example shows that wishful thinking does not necessarily involve a false belief, in the usual sense of that term, i.e., it may be more like fantasy or day-dreaming, and we do not normally say of the day-dreamer that he has a false belief. The wishful thinker accepts \( p \) (e.g., it is the case that he has not won the Irish Sweepstakes) and yet wishes for not-\( p \), i.e., thinks what it would be like if not-\( p \) were true, he knows \( p \) but wishes for not-\( p \) (he does not believe not-\( p \)). The wishful thinker entertains a possibility, the self-deceived believes falsely. Even where the wishful thinker does hold a false belief, it is usually because he does not have all the evidence for \( p \). If he held on to not-\( p \) when faced with the evidence which the rest of us accept as leading to \( p \), he could then be called self-deceived. Wishful thinking, then, may shade into self-deception if, when presented with the evidence which is contrary to his belief, the wishful thinker continues to believe not-\( p \) and denies \( p \). Wishful thinking is usually either a case of not holding a false belief (i.e., something akin to daydreaming) or of holding a false belief but not yet having all the evidence others have which goes against the belief.

Another aspect of wishful thinking is that of its self-ascribing character. We find nothing strange about self-ascribing expressions of the wishful thinking sort, e.g., "I am just engaging in wishful thinking when I think about winning the Irish Sweepstakes." But we do find it strange to use a similar expression in the self-deception case, e.g., it would be rather strange for Mrs. Smith to say, "I am just engaging in self-deception when I believe that Timmy is a good boy." There is a legitimate first person present tense usage of wishful thinking expressions but there is no such use for self-deception. However, there is a place for hypotheticals in self-deception, e.g., "I may be deceiving myself but I believe that Timmy is a good boy," and even a place for expressions like, "I try to convince myself that Timmy does not really mean to hurt the smaller boys." But both of these expressions express doubt as to whether not-\( p \) is true, whereas in self-deception the self-deceived does not doubt not-\( p \) but sincerely says he believes not-\( p \), and acts in accordance with that belief. The expressions which do allow this self-ascribing character to self-deception are of the
sort I have just mentioned, i.e., those which express doubt (others might be those which are said in jest).

III

In our discussion of the epistemic aspect of self-deception, we should examine a few more examples of our everyday talk about self-deception and see if such an examination can shed some light on the role of knowing and believing in self-deception. If we ask the self-deceived person whether he just believes not-\(p\) or knows it, he would usually answer that he knows it (e.g., "I know Timmy is a good boy"). If, on the other hand, we ask him if he knows or believes \(p\) he will deny such knowledge or belief (e.g., "Timmy would not cheat or steal"). It is an important fact about a self-deceived person that he will, apparently sincerely, say he knows (or believes) not-\(p\), but will not say he knows (or believes) \(p\). It is important because part of the way we judge what a person knows or believes is by what he sincerely says, and what the self-deceived sincerely says indicates that he (thinks he) knows not-\(p\) and that he does not know (or believe) \(p\).

In first person present tense usage the self-deceived characterizes his false belief as knowledge but he does not do so when he is no longer self-deceived (i.e., in first person past tense). The first person past tense usage indicates that he sees his former belief just as others see it. It also shows that he did not know \(p\) at the time he was self-deceived. He says, about \(p\), "I just couldn't believe it," or, "I should have known it but I didn't." These utterances seem to indicate that the evidence was there but he did not take it as leading to \(p\), i.e., he did not know \(p\).

How does one come out of self-deception? Not by acknowledging what one (the self-deceived) already knows, but by acknowledging what others already know, i.e., it is when we recognize, when we know, \(p\) that we are finally out of self-deception. When we say that a person is no longer self-deceived because he now acknowledges what he "really knew," what we should be saying is that he now acknowledges what we already knew. The reason we called him "self-deceived" was not because he knew \(p\) and believed not-\(p\) but because he believed not-\(p\) though all others around him believed \(p\), and no other was deceiving him. It is his prior belief in not-\(p\) and his emotional need which is satisfied by this belief, which is active at the time he is self-deceived, which causes the misinterpretation, and thus the self-deception.
IV

We must now try to differentiate between one who is reluctant to believe and one who is self-deceived. The reluctance to believe case is similar to the self-deception case in that the person who is reluctant to believe has a motivated belief, i.e., he has a desire to believe not-\( p \), although \( p \) is the case. Reluctance to believe occurs, for example, when someone who is very close to us dies and we think, and even say, "I know he is dead, but I can't believe it." Expressions like, "I know \( p \), but I can't believe it," "I know \( p \), but I find it hard to believe," "I know \( p \), but I'm reluctant to believe it" are examples of reluctance to believe expressions. Now it is immediately clear that this person, let's call him "the reluctant one," is different from the self-deceived person in an obvious way, i.e., the former will say he knows \( p \) and the latter will not. The reluctant one knows \( p \) and says he knows \( p \) (but is reluctant to believe it) whereas in self-deception the self-deceived person does not say he knows \( p \) nor does he say he is reluctant to believe \( p \), but simply does not believe \( p \). The reluctant one will usually admit that he knows \( p \) and also express his disinclination to believe \( p \). The self-deceived, on the other hand, neither admits he knows \( p \) nor does he express reluctance to believe \( p \). On the contrary, he emphatically denies \( p \).

The self-deceived would not characterize himself as reluctant to believe although others might characterize him that way. Just as there is no first person present tense use for self-deception, so there is no first person present tense use for "reluctant to believe" for the self-deceived about that which he is self-deceived (whereas there certainly is such a use for the reluctant one). Furthermore, the self-deceived person does not see where the evidence points, the reluctant one does. The reluctant one does know and says he knows \( p \).

V

So far I have denied the claim that the self-deceived knows (or believes) \( p \). Distinctions were made between self-deception and cases of ignorance, false belief, mistake, wishful thinking, and reluctance to believe. In the course of my discussion I have included some of our everyday talk about self-deception and shown how our ordinary language applies. However, there are still some formidable questions remaining which have not yet been answered, questions such as the following: If the self-deceived does not know \( p \), why do we call him "self-deceived"? Connected to this question
are issues dealing with other characteristics which we usually associate with self-deception, e.g., intention and act of deception, i.e., what is the role of intention, and what about the act of deception?

I suggest that the role of "self" in "self-deception" is to point out that the person who is deceived is not being deceived by others and is not being deceived by appearances but is somehow the cause of his own deception. In exactly what way he is the cause of his own deception will become clear shortly, but it is not as an agent engaging in an act of deception upon himself. In the "he knows p but persuades himself of not-p" approach, it appears that, in self-deception, the person who is deceived must also be doing the deceiving, i.e., the self, by engaging in an act of deception (e.g., by "persuading himself to believe" not-p), is the agent of deception. But is it ever possible to be deceived without there being an agent who is engaging in an act of deception? I think it is possible: for example, a person of whom we say, "he is being deceived by appearances," i.e., the "appearances are deceiving" case--let us call this "appearance-deception." The appearance-deception case is one where a person may be deceived without an agent committing an act of deception. An example of appearance-deception could be the one of the straight stick which looks bent in water. In appearance-deception, the person is deceived by appearances and no one else has contrived appearances to mislead him.

In the inter-personal deception (other-deception) case we have an agent (deceiver) who engages in an act of deception on the victim (deceived). We can usually point to his act of deception, e.g., the villain, who knows p is true, intending to mislead, tells the victim that not-p is true and the victim is convinced of the truth of not-p. In the self-deceived case, we can readily establish that the self-deceived is deceived but we have trouble ferreting out the

16 See footnote 5 above, especially Demos' description.

17 I take it that an act of deception involves volition, so that even if someone wanted to construe the appearance itself as somehow the agent of deception the volition aspect which is necessary to engage in an act of deception would be lacking. On the other hand, if one could somehow construe the person himself as somehow the agent of the act of deception in the appearance case then we are not talking about appearance-deception but about what is usually called self-deception.
act of deception. What would the self-deceived's act of deception be. It might be argued that it is obvious what the act of deception is, viz., the self-deceived "persuades himself to believe" not-p. I shall discuss this notion later. What is important to realize at this point is that when we talk of deception there are at least two other kinds of deception, i.e., other-deception and appearance-deception. In other-deception there is a deceived, a deceiver, and an act of deception. In appearance-deception there is a deceived but not an act of deception. Obviously, it is not my intent here to assimilate appearance-deception to self-deception but only to show that there are legitimate cases of deception which do not require an agent or an act of deception, and this may help us to break the hold of the other-deception model which could lead us to look for an agent and an act of deception.

The role of "self" in "self-deception" is to point out that the self is the cause of the deception but not that the self is engaging in an act of deception, just as the role of "appearance" in "appearance-deception" (e.g., "he is being deceived by appearances") is to show that appearances are the cause of the deception, but there is not an act of deception being committed. The function of "self" in "self-deception" is to eliminate not only others as the cause of the deception but also to eliminate appearances as the cause of deception. In other words, it is not just that no other is deceiving us, which is what philosophers usually focus on and why they become captives of a certain picture (viz., other-deception), but it is also that appearances are not deceiving us. By bringing out the appearance-deception case we can now look at the self as cause in a different light. Just as appearance-deception means that the person is deceived and it is the appearance which causes his deception, so it is that in self-deception, he is deceived and it is the self which causes the deception. But in both cases there is no deceiver who knows, nor is there an act of deception being perpetrated on the one who is deceived. The self is the cause, not in the sense of the self taking an action to make oneself (persuade oneself) to believe not-p, but, rather, in the sense of the self having a false belief, which he takes as true and which he maintains even to the extent of distorting or rationalizing clearly contrary evidence. If we keep the appearance-deception case in mind as well as the other-deception case we will see that sometimes what a person does not know is what has to be emphasized to explain his deception. Keeping both appearance-deception and other-deception in mind helps to maintain our balance, but when we only have other-deception cases in mind all the weight seems to come down on the side of "he must know."
Sometimes it seems that those who maintain what I call the "he must know" view are not talking about self-deception at all but about someone who knows p, is reluctant to believe p, and wants to believe not-p, i.e., the reluctant one. However, if accused of this mistake, the rejoinder would be that although it is true that their picture of the self-deceived does allow for him to be portrayed that way, it is only if one takes "he persuades himself" as meaning he attempts to persuade himself but is not quite successful, in which case he wants to, but does not quite, believe not-p. "However," they could add, "we are not saying he tries to persuade himself and fails, but that he succeeds in persuading himself." This rejoinder brings us face to face once again with what may be our real "linguistic snake-in-the-grass" here, viz., "he persuades himself."

The first thing we should note about "he persuades himself to believe not-p" (or perhaps, an expression which is also quite common in our ordinary talk of self-deception, "he convinces himself") is that just as self-deception is modeled on other deception, so self-persuasion is modeled on other-persuasion. Let us look at a case of B (the person to be persuaded) as a person who knows p. Along comes A to persuade B of not-p. If A's persuasion is successful then B will drop his belief in p and now believe not-p, i.e., even if what B formerly believed was true, once A has persuaded him that not-p is true, B no longer believes p. If he still believes p then obviously the persuasion has failed and B has not been persuaded (or convinced) of not-p. In applying this view of persuading or convincing another to self-deception we can see that if A persuades (or convinces) himself of not-p then (1) he no longer believes (or knows) p or (2) he has not persuaded himself. If (1), i.e., if he does not know p,

18 Presumably, proponents of this view are not just using "he persuades himself of not-p" as being equivalent to "he deceives himself," but rather that the former comes first and the latter follows from it., i.e., A knows p and after persuading himself to believe not-p he enters a state of self-deception. In any case, there certainly is a distinction to be made between persuading and deceiving. I may persuade you that what you formerly believed is wrong and if what you formerly believed is, in fact, wrong then I have persuaded you, not deceived you.
then according to the "he must know" view, he is not self-deceived. If (2) holds then a crucial element of that view no longer pertains, viz., he did not persuade himself of not-p. But if he is not persuaded of not-p then all we are left with is that he knows (and believes) p and does not believe not-p. Which means that self-deception does not pertain.

VII

Hovering in the background of the discussion so far is the shadow of intention in self-deception, i.e., can we talk of self-deception without bringing in intent to deceive? This question must now be confronted. The argument for intention is that the self-deceived intentionally deceives himself, i.e., he persuades himself to believe the opposite of what he knows to be the case. This view takes as its fundamental assumption the notion that the self-deceived knows p. His intention, I want to maintain, is not to deceive himself (he does not say or think, "now I have to fool myself") but to put not-p in the best possible light. He selects and interprets the facts to suit himself. The result may be that he is deceived (about p) but his intention was to preserve his initial belief (about not-p), which he took as true. My view of his intention is not based on an unproven assumption but on what we (as observers of the self-deceived) do know, viz., the self-deceived believes not-p.

Some philosophers talk as if the self-deceived sets out to deceive himself, i.e., he knows what is the case but "purposefully" deceives himself, he "persuades himself" to believe the opposite of what he knows to be true. In the only book written by a philosopher on self-deception we find this:

The crucial element which is necessary... is the element of purposefulness. If our subject persuades himself to believe contrary to the evidence in order to evade, somehow, the unpleasant truth to which he has already seen that the evidence points, then and only then is he clearly a self-deceiver.... It is correct to say that, even if we assume no motive, we would call a person self-deceived if he persuaded himself to believe what in his heart he knows is not so.19

19 H. Fingarette, Self-Deception, page 28 (italics in the original).
In my view it is not that he "purposefully" (perversely? maliciously?) deceives himself, but that he believes falsely and then "purposefully" (mis)interprets the evidence to maintain what (he thinks) he knows.

But, it might be argued, "purposefully" suggests intent, in this case, the intent to deceive. The intent of the self-deceived person is to support what he believes to be true. He purposefully goes about trying to justify his belief, i.e., purposefulness is certainly there, but the purpose is to maintain what he thinks is a true belief. He does not say to himself (if we can talk about what he says to himself at all), "even though p is the case, I will ignore it and try to show that not-p is the case." He is more likely to say, "Although others think p is the case, I know otherwise." Here he admits that others have a different view than he does, but he thinks that their view is wrong, because he believes (thinks he knows) otherwise, he then proceeds to misconstrue the evidence and/or bring in some of the reasons upon which he formed his initial (false) belief. In other words, there is intention involved, but it is not intent to deceive (i.e., to make oneself believe not-p although one knows p) but intent to justify what he believes to be true (not-p).

It is true that the self-deceived is deceived in that he believes not-p. It is also true that he is deceiving himself, not in the sense that he knows p and engages in an act of deception, but, rather, that through his misinterpretation of the evidence (due to his false belief and the need this belief fills for him) he is the cause of his own deception. So if we want to talk about an act in self-deception, it is his act (or acts) of interpretation, of making a mistaken inference from the evidence, coupled with his false belief, which leads him into the misinterpretation, and, hence his being deceived about what is the case. Through his intent to justify what he considers a true belief, he unintentionally deceives himself. So he, unintentionally, is the cause of his own deception, he is self-deceived. The cause of his self-deception is not an act of deception (an intent to deceive) but an act of justification, i.e., an attempt to justify what he considers true. The effect of his act is that he, unintentionally, deceives himself. Prior to his misinterpretation of the evidence, he just held a false belief, but once that false belief is perpetuated by attempts to justify it (to himself and/or others) he becomes self-deceived. His misinterpretation of the evidence and his reaching the conclusion not-p shows that he is deceived. The effect of his misinterpretation of the evidence is that he continues to believe, contrary to the evidence, in not-p, and is now self-deceived.
Although it is unintentional, he has now become the cause of his own deception. But notice that, qua agent, he is not an agent of an act of deception (there is no question of his "purposefully" attempting to deceive, he does not "persuade himself") but his act is one of attempted justification (of what he thinks is a true belief), and the result of that act is that he causes, not his false belief (which would appear to be the case if we assume that he knows p but persuades himself of not-p), that was already there, but the continuance of his false belief despite the contrary evidence. Now, rather than characterizing him as just ignorant or mistaken, we may characterize him as self-deceived, because it is now his own actions which cause him to be deceived, whereas before he might have been deceived by others, been ignorant, etc.

Why does the self-deceived person misinterpret the evidence? It is not that he knows p and perversely argues for not-p, but that he thinks (wrongly) that not-p is true and wants to justify that belief. He knows what the evidence is which, for others, leads to the conclusion p, but he does not believe that the evidence forces the accepted conclusion on him. His attempts to justify his false belief lead him to make a mistake at the level of inference. Where we would go straight ahead to p he makes a swerve and comes up with not-p. We, naturally, find his move unsatisfactory, although he seems satisfied with it (like Lucretius' "swerve," it fits his view, but leaves the rest of us frustrated by his move.)

What makes the "he must know" view want to say the self-deceived misinterprets the evidence and thus does not reach the conclusion p and still say he knows p? They want to say that this misinterpretation of the evidence (without the false belief, since up to this point, on this view, he knows p., i.e., he has a true belief) shows that he is self-deceived. Let us look at what the misinterpretation of the evidence alone (without the false belief and the attempt to justify that belief which I claim) does show. It shows: (1) He has misinterpreted the evidence. (2) he reaches a false conclusion. Se we can say that it shows that he does not know p. But that is what he has been saying all the time. Or do we want to say that he first interprets the evidence correctly (he knows p) and then, on a second reading as it were, he misreads the evidence, i.e., he persuades himself of not-p. This just brings us back to the intention argument and that, as I have shown, needs more arguing.
We now have a new picture of self-deception. It is a picture of a person who is deceived in that he does not know p. A person who becomes, unintentionally, the cause of his own deception by attempting to justify a belief (in not-p) which he thinks is true, but which, in fact, is false. His belief in not-p is a desirable belief (something which gratifies a need or desire of his) and his attempt to justify this belief leads him to misinterpret the evidence and to continue in his false belief over an extended period of time. There is no act of "persuading himself to believe" not-p, nor is there any intent to deceive himself.

If the above analysis of self-deception is correct, then the shadow of paradox which we noted at the beginning of the paper has been effectively removed and the phenomenon emerges in a clearer light.

Dante A. Cosentino