

Kent Bach:
(Apparent) Paradoxes of Self-Deception and Decision

§ I. What Self-Deception Is (Not)

There are two basic questions about self-deception itself, which take familiar philosophical forms:

___ **What is it?**

___ **How is it possible?**

Davidson:

___ **One is intentionally self-deceiving.**

Bach and Mele:

___ **There is no need to regard self-deception as intentional just because it is motivated and purposeful.**

Bach's claim:

___ **Much of the mystery surrounding self-deception is lifted once we abandon the assumption that the self-deceiver must be acting intentionally if he is to be acting purposefully.**

§ Suppression

Q: But what *is* he doing? What does he accomplish and how does he manage to accomplish it?

[the Orthodox View]:

___ **The subject gets himself to form a contrary belief. He does not change his mind, in the sense of replacing one belief with a contrary one, but adds the contrary one to his stock of beliefs.**

[Bach's Proposal]:

___ **Rather than adopt a new, contrary belief (as on the orthodox view), what the self-deceiver does is to keep himself, at least on a sustained and recurrent basis, from thinking what he believes. No contrary belief is needed to suppress or inhibit the effect that the unpleasant belief normally has on this thinking, although the self-deceiver may need to clutter his mind with reasons against the unpleasant belief and with thoughts to the contrary.**

One effect of self-deception, then, is to suppress the subjective evidence that one believes that *p*, hence to make it seem to one that *not-p*.

§ Methods of Suppression (not to think what one believes)

(i) Rationalization

Rationalization can involve distorted weighing of reasons or evidence and even fabrication of plausible but phony reasons (or motives, when one's own actions are involved).

(ii) Evasion

One avoids the thought *that p* by avoiding the thought *of p*.

(iii) Jamming

One clutters one's mind with thoughts contrary to the unpleasant belief or contrary to evidence one has in support of that belief.

§ Exclusionary Categories and Selective Attention

<selective attention>

Given our limited attentional and cognitive resources, we must be selective in what we consider in a given situation.

<exclusionary categories>

We each possess an arsenal of "exclusionary categories" that we apply to topics, doctrines, actions, and persons (among other things) in order to justify not taking them into account in our thinking.

Examples: *absurd, dangerous, crazy, impossible, irrelevant, offensive...*

Q: What is the role of exclusionary categories in self-deception?

— If something is too painful to consider, "too hard to deal with", one has a practical reason for not considering it.

The use of exclusionary categories is not inherently irrational. Of course, the misapplication of exclusionary categories can deprive one of opportunities to correct errors in one's thinking and to entertain new possibilities.

* Bach's Analysis of Self-deception:

There is a cleavage between what one believes and what one occurrently thinks. This cleavage is possible because, contrary to Descartes' dictum that "there is nothing in my mind of which I am not in any way conscious," beliefs do not have to be conscious to be held. Recognizing the cleavage and why it is possible removes the semblance of paradox in self-deception.

§ II. Paradoxes in Decision Theory

* Newcomb's Problem

There are two boxes, the transparent Box A, which contains \$1000, and Box B, which may or may not contain \$1,000,000, and you can take either the contents (if any) of Box B alone or the contents of both. You are told that a highly reliable predictor has placed \$1,000,000 in Box B if, and only if, he predicted that you would not choose both boxes, and he knows that you are told this. What should you do?

Choose both → end up getting 1,000

Choose only B → end up getting 1,000,000

Not choose both → there is 1,000,000 in box B

___ It reflects a tension between two kinds of decision theory. Two-boxers generally rely on causal decision theory, which discounts considerations that do not make a causal difference. They reject the one-boxer's appeal to the counterfactual conditional. The one-boxer urges you to remind yourself of the predictor's past performance.

[I] the Causal Decision Theory

___ Only the future is causally affected by our present decision. The past is gone.

[II] Evidentialism

___ The evidentialist is claiming that there is a dependence between the action and the prior prediction – but not a causal dependence. One's action provides evidence for the prior prediction of it.

This noncausal dependence is described by the following two counterfactual conditionals:

[A] If I were to choose one box (probably) the predictor would have predicted that I would choose one box.

[B] If I were to choose two boxes (probably) the predictor would have predicted that I would choose two boxes.

In the theory of counterfactual conditionals these are called “back-tracking” conditionals. Causal theorists disallow them out of hand, but evidential theorists do not.

*** Alvin Goldman’s Book of Life**

In this case, there is no issue of changing the past. Rather, the issue concerns one’s access to information about the past. The peculiarity is that one cannot possess the relevant information while being in a position to act on it.

§ Bach’s Critique of Dupuy’s Dualities

Dupuy’s discussion is needlessly complicated by a number of philosophical red herrings:

- (1) changing the past
- (2) determinism
- (3) two kinds of temporality
- (4) two kinds of rationality

(1) Changing the Past

Dupuy’s discussion of Calvinism misleadingly suggests that what the believer does is to try to change the past.

When you perform an action, the world is different from how it otherwise would have been. However, this does not mean that you have *changed* the future; you do not change how it *will* be but make it different from how it otherwise *would have been*.

(2) determinism

Dupuy asserts that the Calvinists must be compatibilists. But their reasoning does not assume the compatibility of free will with determinism. What it does assume is divine predestination.

Nor does the Calvinists’ predicament depend on their having free will.

In any case, determinism is not the issue. For what the Calvinist (or the one-boxer) believes is that his action is determined by the same thing, whatever it is, that determines his destiny. He does not know what that thing is and he does not know what his destiny is – except by how he acts. Thus he can act in such a way as to provide himself with evidence both for what determined his destiny and for what his destiny is, but he is shielded from knowing what this is before he acts.

(3) two kinds of temporality: “occurring” and “projected”

However, temporality appears to be inessential to the issue. Essentially the same problem can arise in a situation where the two players must act at the same time.

(4) two kinds of rationality: two principles of rationality – causal and evidential

Granted, there are two competing principles of rationality, which underlie causal and evidential decision theory respectively, but it is questionable whether they are both right.

§ Bach's Critique of Dupuy's Solution

What Dupuy describes as a counterfactual power over the past is a subjective illusion.

Given determinism and an initial state of the universe, whatever does happen must happen and nothing else could happen. But since one is ignorant of the details, one has the illusion that performing an action does not lead to what was bound to happen but somehow realizes one possibility and forecloses others. For as long as one is ignorant of what is bound to happen or of its causes, one is free—epistemically free—to provide evidence for facts that are otherwise inaccessible.

One is not metaphysically free to do this because what one does (assuming determinism) is itself part of what was bound to happen.

§ Bach's Critique of the Artificiality of Decision Theory

Some common assumptions of decision theory:

1. It is assumed that a given problem is explicitly posed to a person.
2. It is assumed that a set of possible choices and outcomes is provided, complete with their expected utilities.
3. It is assumed that the relevant information is available to the person, and presented as relevant.
4. It is assumed that there is no time limit on the decision.
5. It is assumed that the decision cost is negligible.

All five of these assumptions are unrealistic, not just collectively but even individually.

§ Bach's Analysis of the (Seeming) Paradoxes in Decision Theory

With the problem of what to do next in mind, we can see the deeper significance of the fact that people are satisficers rather than maximizers. We satisfice not so much

because we are content with less than the best, and not merely because we operate under obvious cognitive and temporal constraints. The real reason is that we have to do something at every moment.

Thus, something's seeming good enough to do at a given moment consists, practically speaking, simply in the unoverridden thought to do it, then and there.

§ Bach's Conclusion

It is clear that the rationality of commitment is paradoxical from the standpoint of a particular point in time.

What is most striking about the sources of the semblance of paradox both in self-deception and in the various decision problems is that they all involve some sort of gap between what one is *at a moment* and what one is *over time*.