Deceived by metaphor

John A. Barnden

Abstract: The views of self-deception that Mele attacks are thoroughly metaphorical, and should never have purported to imply the existence of real internal acts of deception. Research on self-deception, including Mele's appealing account, could be enriched and constrained by a broader investigation of the prevalent use of metaphor in thinking and talking about the mind.

Self-deception has interesting connections to metaphors of mind that are commonly used in everyday discourse. Papers on self-deception often use colorful metaphors, but such authors do not commonly *mention* the metaphors as such. I will suggest that the views of self-deception that Mele attacks are thoroughly metaphorical conceptions. Even though these conceptions may be pragmatically useful in common-sense thought and discourse about self-deception, it may well be that what is really occurring during self-deception is mental processing of the style claimed by Mele. Whether Mele is right or not, further study of self-deception would be enhanced by being placed in the context of an investigation of the use of metaphor in thinking and talking about the mind. (I have been engaged in such an investigation as part of an artificial intelligence research project – see Barnden [1996] and Barnden et al. [1995].)

Ordinary speakers and writers frequently refer to parts of the minds of (mentally healthy) people as if those parts were themselves persons, with their own mental states, emotions, and so forth, and often engaging in natural language utterances. Here are some examples (from real discourse, with minor adaptations):

(1) "One part of Mike knows that Sally has left for good."

(2) "Part of Mike was insisting that Sally had left for good."

(3) "Part of you wants to talk about your personal problem but part of you hates the idea."

(4) "Half of me whispered that I'd drive all the way there."

(5) "It was as if his consciousness didn't want him to be without anxieties."

(6) "Did part of you think, 'Yes, I'm flattered'?"

I view such discourse chunks as manifestations of a conceptual metaphor, "mind parts as persons" (MPaP). I would also claim that the metaphor is manifested in at least some readings of sentences such as

(7) "Sally told herself that Peter was faithful."

In many contexts as appealing, partial paraphrase would be that "one part of Sally was trying to convince another part that Peter was faithful."

In manifestations of the MPaP metaphor, there is generally a strong connotation that mind parts that are *not* mentioned do *not* have the mentioned belief or desire. In (1) we should presumably take it that (metaphorically speaking) some other part of Mike does not know that Sally has left him for good. And often it is reasonable to take the unmentioned parts actually to have a contrary belief, desire, and so on. In (2) the use of "insisting" strongly suggests that (metaphorically speaking) some other part of Mike has claimed, and believes, that Sally had not left for good.

Now, the interpersonal view of self-deception that Mele attacks models self-deception on ordinary deception by one person of another. One common elaboration of this view is that the self-deceiving person contains two subsystems, one of which intentionally deceives the other into believing something. This elaborated view could, conceivably, be either a literal one or a metaphorical one, where of course a prime candidate for the metaphor is MPaP. However, the literature on self-deception rarely mentions metaphor explicitly (but see some exceptions herewith), and discussions appear generally to assume without comment that the multiple-subsystems view is to be taken literally.

On the contrary, I suggest that the notion of self-deception, as it appears in stereotypical cases, is inherently metaphorical, and involves entirely metaphorical stances such as that one sub-person consciously intends to deceive another sub-person. It could well be useful and economical for us to adopt such a metaphorical view in everyday thought and discourse about self-deceivers. Nevertheless, the practical convenience of the metaphorical view does not imply that the objective, scientific truth of the matter is that the self-deceiver contains subsystems corresponding to the metaphorical sub-persons, or that, even if there are such sub-systems, that they do anything that could literally be called entertaining beliefs and intentions or engaging in acts of deceit, any more than the metaphor of death as a person implies that death really has beliefs and intentions. Therefore, the way is open for the mind to be operating in the way that Mele suggests it does in his appealing account.

Consider the question of what scientific sense can be made of "a part of" Mike believing something *P,* or "insisting" something, and so forth. In the case of such statements, where the use of metaphor is relatively blatant, it should not be considered shocking to claim that neither "Mike-as-a-whole" nor any identifiable sub-system within Mike can *literally* be viewed as believing *P,* insisting *P,* and so forth. Rather, we arguably have at most the right to say that *in some sense* Mike-as-a-whole believes *P.* And there is no contradiction in saying that Mike-as-a-whole in some sense believes *P* and in some sense believes *not-P,* because the "senses" could be different; we have no warrant to conclude that Mike-as-a-whole believes *P-and-not-P* in any sense.

Then, if I am right that common-sense notions of self-deception are inherently metaphorical, a scientific account of what really underlies self-deception should be continuous with a scientific account of what is going on behind garden-variety statements such as "a part of Mike believes." That being the case, there should be little impulse to suppose in the first place that self-deceit *really* involves a contradictory state of mind or any real intention by the agent or a subsystem of the agent to deceive anyone or anything. I therefore suggest that the study of self-deception could be broadened, enriched, and constrained by considering the relationships of that notion to other metaphorically described mental states and processes.

MPaP is not the only relevant metaphor. In some accounts of self-deception, for example, that of Davidson (1985) alluded to in Note 5 in the target article, the self-deceiver's mind is viewed as partitioned into several regions, where the boundaries between regions cannot be crossed in some relevant sense: for example, the contradiction between *P* in one region and not-*P* in another cannot

be seen by the person. These accounts rely on the metaphor of "mind as physical space," another extremely prevalent metaphor in ordinary discourse.

 Finally, metaphor and related matters do occasionally receive mention in the literature on self-deception. For instance, Johnston (1988, p. 82) briefly mentions the tack of taking a subsystem account of self-deception metaphorically, but does not pursue the matter. Rorty (1988) bases an account on a superimposition of two "pictures" of the mind, and Bittner (1988, p. 538) casts talk of quasi-human parts of a self-deceiver as a "myth." Perhaps those authors would be happy to take the pictures and the myth, respectively, to be metaphors. Mele (1987, p. 3) quotes another self-deception researcher (King-Farlow 1963) as claiming that a person can quite often usefully be "looked at" as a large, loose sort of committee. Here King-Farlow is close to talking explicitly about MPaP.