

Metaphor, Self-Reflection and the Nature of Mind

(extended abstract of proposed chapter for
“Visions of Mind: Architectures for Cognition and Affect”)

John A. Barnden,
School of Computer Science,
University of Birmingham, B15 2TT, United Kingdom

J.A.Barnden@cs.bham.ac.uk
Tel: (+44)(0)121-414-3816 Fax: (+44)(0)121-414-4281

1 Introduction: What Questions are we Addressing?

This chapter addresses aspects of the following central issues for the book:

- a- What is mind?
- b- What are theories of mind?
- c- Nature of computationally implemented architectures and systems based on theories of mind
- d- Fragmentation in the study of mind.

But it does so largely from the point of view of *how does a mind view itself or other minds* rather than from the theoretical observer’s point of view of determining what the mind *really* is, in any direct sense anyway. As a result, the chapter’s take on issues (a-d) is roughly as follows, where (a) and (b) have been collapsed together:

- ab’- How does a mind view itself (what theories does it have about itself), how does it view other minds, and how do these matters interact with the question of what minds *really* are?
- c’- Nature of computationally implemented architectures and systems involving minds’ views of minds.
- d’- Fragmentation in minds’ views of minds.

The move to these issues from issues (a-d) might be thought to be twisting the latter too far. But behind (ab') is a claim that how a mind views itself is part of, and can affect, the real nature of mind itself. The most basic point here is that, since an important aspect of mind is the process of its thinking (consciously or unconsciously) about itself, the fact that it views itself in a certain way just is fact about that mind, no matter how inaccurate the view that it has of itself is. To give an extreme example to make the point, if a mind thought it was made out of fire and water, then the fact that it did so is an important fact about that mind, even though it was not actually made out of fire and water. To put the point another way, theories of mind must take into account the views and theories that minds have about themselves and each other.

But, more deeply, the chapter will claim that a view that a mind has of its own nature can, so to speak, entrain that mind to become more in accordance with that view than it would have been without the operation of the view. In essence, and to a partial degree, views of oneself can become self-fulfilling prophecies.

The main ideas in the chapter are to do with an individual mind's view of itself, rather than of other minds, and is thus consonant with the need to include self-reflection capabilities in architectures of complete minds (cf. the self-reflective, "meta-management" layer in the mental architecture that Aaron Sloman has proposed). However, the chapter also notes that a view that a mind has of itself could be influenced by the views it perceives other minds as having of themselves and each other. Views of mind, just as of anything else, can be transmitted from person to person.

The chapter implicitly contributes a little on counteracting the fragmentation noted in issue (d), in a few different ways: it brings a computational outlook to bear on deep philosophical and psychological issues; it naturally supports a link between the study of language about the mind and the study of mind itself (see next section); and, finally, concentrating on self-reflection has a natural integrative effect on one's thinking about mind, since self-reflection involves reflection about many different aspects of the self that is reflected upon. On the other hand, the chapter's specifics are really more about (d'), and are indeed in the direction of proposing that the mind has a natural tendency to have a fragmented overall view of itself.

The chapter takes "mind" to include affect. The way a mind views its own affect is important here, as are the affective aspects of the way a mind views itself (even the non-affective features of itself).

2 A Specialized Slant: Metaphor-Based Self-Reflection and Self-Management

One aspect of a complete mind, situated within anything like our world, must be the ability to reason about other minds and about itself as a complete mind. Now, as cognitive linguists and others have shown (see, e.g., Lakoff 1993), much human discourse concerning minds is highly metaphorical. For example, we commonly talk about each others' minds as being physical containers, of each other as being made up of competing sub-people, of ideas as living creatures or as inert physical objects, of understanding as if it were physical perception, etc. We sometimes using one metaphorical view and sometimes another, possibly

conflicting one. Sometimes we mix different views together. This is all in perfectly mundane discourse, not (just) poetry and other literary art.

A further tenet held by many metaphor researchers (e.g. Gibbs 1994, Lakoff 1993) is that the metaphorical views used in discourse are, generally speaking, crucial aids in thought (conscious or unconscious) rather than just linguistic icing. Assuming that this is true, people's conscious and unconscious thinking, not just their discourse, about each other and about themselves is partly, and perhaps highly, metaphorical. It is therefore plausible to suggest that natural minds do, and artificial complete minds should, reason about each other and themselves in metaphorical terms. This is not to say that minds *believe* that the views are true—they are just useful ways of thinking.

Moreover, suppose a metaphorical view that someone takes of some entity can affect not just the person's reasoning and communication about it but also how the person deals with it (interacts with it, manipulates it, controls it, etc.). For example, thinking of poverty as a disease can affect one's attempts to control it. Similarly, metaphorical views that a mind has of itself could affect its self-management operations, not just its reasoning about itself. For instance, if it is thinking of its own ideas as living, sentient creatures, it may take a more passive stance to dealing with those ideas, on the assumption that they will move and interact of their own accord.

The chapter concentrates entirely on metaphorical aspects of self-reflection and self-management. However, some of the issues could be generalized to become observations about self-reflection and self-management in general.

In natural language discourse, affective states are often described metaphorically (see, e.g., Fainsilber & Ortony 1987, Kövecses 2000). A mind's reflection on its own affect can therefore be conjectured to involve metaphor. Also, metaphor is often used in natural language discourse to convey value judgments and emotions about the targeted subject matter (often to deviously smuggle them in, but also, more beneficently, to convey them in an economical and effective way). For instance, thinking of poverty as a disease could cause one to have particular negative emotions about poverty or indeed poor people. Thus, we may conjecture that affect could be held within minds in a partially metaphorical way.

3 Metaphorical Self-Reflection as a Practical Necessity

The previous section mentioned metaphorical self-reflection as a mere possibility—just one way in which minds might think about and manage themselves. However, it is plausible to suggest that there may in fact be no practical alternative to doing substantial amounts of reasoning about mental states by means of metaphor.

First, as can be seen from the literature on metaphor, it is plausible that there is no practical alternative

to metaphor for thinking about messy abstract domains, especially when matters are complex or subtle. It is widely acknowledged that metaphor, when applied appropriately to messy domains, can provide more economical and precise description, and more effective reasoning, than is otherwise practical (or perhaps possible). Secondly, it is plausible that one's own mind is a messy/complex/subtle domain for oneself as well as for others, even if one has some sort of privileged, direct access to one's own mind. Indeed, the potential messiness, complexity and subtlety is, if anything, increased by having more extensive access to one's own mental states than to those of other minds.

Furthermore, an agent X can be expected to learn metaphorical ways of talking and thinking about minds, from the metaphorical ways that other agents use in their speech. These ways of talking about minds could be absorbed by X and become ways in which X thinks about itself. This does not, of course, preclude X developing metaphorical and other ways of thinking about itself purely through self-reflection.

4 Distorting Oneself through Metaphor

We pointed out above that a metaphorical view used in a mind's self-reflection is a real feature of the agent, no matter how inaccurate the view itself is. In that sense, the claim is that metaphorical self-views are aspects of the real nature of mind. But we can also see ways in which the use of a metaphorical self-view can cause the agent to become more similar to its own view of itself than it would otherwise have been. Views of oneself can become self-fulfilling prophecies, to some degree. We will look at two possible ways in which this could happen.

4.1 Distortion Method 1: Through Metaphorical Self-Management

Any given, broad metaphorical view of mental states and processes (such as the view that the mind is a physical space and ideas are inert physical objects located in moving around inside) captures some real aspects of mind and ignores others. This is just a special case of a general feature of metaphorical description—different metaphorical views generally capture different aspects of what is being viewed (Grady 1997; Lakoff & Johnson 1980). Moreover, even when a view captures certain features, it generally does so only approximately.

Thus, if a mind's self-management is partially influenced at some point in time by a particular metaphorical view V, the self-management may be partially defective because of the inaccuracies of V. But, the operations of self-management may themselves to some extent tend to make the mind behave as if it were indeed more accurately described by V. We can call this phenomenon distortion of oneself through metaphorical self-management.

One way this could happen is if the metaphorical view fails to be sensitive to particular opportunities

for external or internal actions by the mind; then, self-management may be deprived of the opportunity for exploiting those possibilities, so that the mind does not perform actions that it could, in fact, perform.

As an example, person Z may be viewing her own current mental operations via the metaphorical view of MIND AS PHYSICAL SPACE. Some of her self-management could be influenced by her perceptions of how “central” some ideas are in her mind-space (that is, her perceptions of how strongly she is attending to them), and could be conducted with the aim of “moving” ideas closer or further from the centre. She might assume that bigger “movements” require more effort, so that less central ideas require more time and effort. She might *therefore* attend even less to ideas she perceives as being on the periphery of her mind, irrespective of whether they are actually significantly more time or effort to deal with or not. Those ideas could then indeed become less attended to than they were already. Thus, the metaphorical view Z is taking of herself can tend to exaggerate certain features of her mental state that are (inaccurately) captured by the view.

A more vivid example is someone who views himself as having an “inner child” partially governing his thoughts and actions. To the extent that the person believes that children should not be overly controlled, or cannot be controlled, he may refrain from taking self-control actions that he would otherwise take (and be perfectly able to take), and thus would come to act and think more as if he really did have a child inside controlling things.

Of course, these observations are just special cases of the more general observation that people are limited by the views that they hold about themselves. Someone who believes they cannot do something will tend not to try to do it, whatever it is.

4.2 Distortion Method 2: Through Metaphorical Self-Conformity

People tend to some extent to adapt to—in the sense of coming to conform to—views other people have of them or ways other people have of dealing with them. For example, if A thinks B thinks A is stupid, A can start to act more stupidly than he would otherwise. Another type of case is illustrated by a situation where a person A acts in a business-like way in dealings with B because B is acting in a business-like way with A. Quite apart from such questions as A’s thinking that he *should* act in a business-like way, in order, say, to impress or outwit B, there is simply the effect that B’s business-like dealing with A sets up a context of action where some types of action are more appropriate than others, and A may simply slide naturally into that context.

Given that metaphorical views that people entertain about each other can affect the way they behave towards each other, it follows that someone may tend to conform, temporarily at least, to some view that is affecting the way someone else is dealing with him.

Could these types of effect apply also within a single agent? Suppose an agent can legitimately be

viewed as being composed of two or more sub-agents, with mind-like capabilities and able to perform operations on each other, reason about each other and communicate with each other. (Even if the human mind is not normally, or ever, like this, it could be the way an artificial agent is organized.) Then is it too fanciful to suppose that a sub-agent A could conform to the way it is being dealt with by another sub-agent B, and therefore in particular come to behave more in accordance with some metaphorical view that B is entertaining about A?

If this could happen, it would be another way in which the overall agent is distorting itself to conform to its own metaphorical self-reflection.

5 Metaphorical Qualia

A rather different line of thought is also suggested by metaphors of mind. First-person manifestations of metaphors of mind are common. Examples such as the following are common in ordinary discourse:

one part of me wants to ...

it was in the back of my mind

the thought crept into my mind

the thought stuck to me

I said to myself that ...

my mind felt totally focused

Perhaps we do not use such language merely because it supports useful reasoning about the described mental states, but also because—at least to a limited extent and for some of the time—it reflects the *feel* of mental states to us, using the word “feel” in a sense as broad as the word “qualia”. (Thus, in the intended broad sense, redness has a “feel”). Using this broad sense, the conjecture is that, for instance,

thought can *feel* like internal speech

thought can *feel* like vision

one’s mind can *feel* like a physical space, and one can *feel* that one’s ideas are far apart or moving around within that space, or coming into the space from outside.

and so forth. Now, if this is the case, then, since these feels are themselves part of the conscious mind, it follows at least some metaphorical views of mind are, in part, aspects of the real nature of the consciousness, not just arbitrary descriptions of mental states.

6 Fragmentation of a Mind's Overall View of Itself

As we stated above, a given metaphorical view captures only part of the targeted phenomenon, and different views target different parts (in general). Also, because of inaccuracies in the capturing performed by different views, the views can conflict in what they convey about the target.

Thus, it is natural to expect that if self-reflection in minds is importantly metaphorical, there will necessarily be an important degree of fragmentation and inconsistency in self-reflection.

However, if individual metaphorical self-views can fulfill their own prophecies (see the above sections on distortion), then it is advantageous to have a variety of distinctly different metaphorical views engaged in self-reflection/management, rather than just one.

The mentioned fragmentation and inconsistency is primarily an observation about human minds. However, if it is true that metaphor provides to artificial minds a useful tool for description of mental states, then metaphorical self-reflection/management could be useful. But this then would bring with the fragmentation and inconsistency. This chapter proposes that this outcome should simply be embraced. After all, non-metaphorical self-reflection would probably have to involve over-simplifications and therefore inaccuracies, and therefore there might need to be multiple, partially inconsistent self-views even if they were all non-metaphorical.

7 A Relevant Implemented System

The author has developed an implemented AI system called ATT-Meta for conducting metaphor-based reasoning (Barnden 1998, 2001, Barnden *et al.* 1994, Lee & Barnden 2001). This has in fact been applied largely to the special case of metaphor-based reasoning about mental states. For example, it can trace through implications of two ideas being “far apart” in a mind considered as a physical region. The intended ultimate purpose of the methods used in the system is for them to form part of natural language discourse processing. However, the techniques used in the system could also be used reflectively by a mind to reason about itself on the basis of metaphorical self-reflection.

(The full chapter will contain a fairly detailed description of the system, in order to demonstrate how the types of mental processing discussed in previous sections could realistically form part of a mind design.)

8 Conclusion

We have focused on the question of the views that minds can have of themselves, as opposed to the question of what minds are really like. However, we have pointed out that these views are themselves part of the

real nature of mind. Moreover, entertaining a particular view of itself can cause a mind to become more like what the view dictates. Although these points apply to any sort of view, we have concentrated on the special case of metaphorical views. Such views may be needed in practical self-reflection, just as they are needed in practical natural language discourse about mind, because of the messiness, complexity and subtlety of mental states and processes. Metaphorical views throw into especially sharp relief the likely partiality and inaccuracy of individual views, and inconsistency between different views. Our work on the ATT-Meta system for metaphorical reasoning provides ideas on how both natural and artificial minds could think metaphorically, and thus make the general considerations of this chapter more real for mind researchers.

Acknowledgments

This research was supported by grant GR/M64208 from the Engineering and Physical Sciences Research Council of the UK. The proposed chapter is developed from a talk created for Aaron Sloman's symposium *How to Design a Functioning Mind* at AISB-00, held at the University of Birmingham. The talk was given in adapted form at the EURESCO conference *Mind, Language and Metaphor: Euroconference on Consciousness and the Imagination*, Kerkrade, The Netherlands, in April 2002.

References

- Barnden, J.A. (1998). Combining uncertain belief reasoning and uncertain metaphor-based reasoning. In *Procs. Twentieth Annual Meeting of the Cognitive Science Society*, pp.114–119. Mahwah, N.J.: Lawrence Erlbaum.
- Barnden, J.A. (2001). Uncertainty and conflict handling in the ATT-Meta context-based system for metaphorical reasoning. In V. Akman, P. Bouquet, R. Thomason & R.A. Young (Eds), *Procs. Third International Conference on Modeling and Using Context*, pp.15–29. Lecture Notes in Artificial Intelligence, Vol. 2116. Berlin: Springer.
- Barnden, J.A., Helmreich, S., Iverson, E. & Stein, G.C. (1994). An integrated implementation of simulative, uncertain and metaphorical reasoning about mental states. In J. Doyle, E. Sandewall & P. Torasso (Eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference*, pp.27–38. San Mateo, CA: Morgan Kaufmann.
- Fainsilber, L. & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2 (4), 239–250.

- Gibbs, R.W., Jr. (1994). *Poetics of mind: Figurative thought, language and understanding*. Cambridge, UK and New York, USA: Cambridge University Press.
- Grady, J.E. (1997). THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics*, 8(4), pp.267–290.
- Kövecses, Z. (2000). *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge University Press.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, 2nd ed. Cambridge, UK: Cambridge University Press.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lee, M.G., & Barnden, J.A. (2001). Reasoning about mixed metaphors with an implemented AI system. *Metaphor and Symbol*, 16(1&2), pp.29–42.