

# Reconciling the Two Images

**Andrew Brook**

Philosophy and Interdisciplinary Studies  
Carleton University  
Ottawa, ON., Canada K1S 5B6  
andrew\_brook@carleton.ca

## Abstract

Some of the indecisiveness and lack of common purpose that characterizes cognitive science at the moment, I argue, can be explained by its lack of success so far in connecting the scientific, computational image (better, images) of cognition developed in cognitive science to people as we experience them in ordinary life, society, law, literature, etc. Following Sellars (1963), we call these two ways of representing cognizers the scientific image and the manifest image. The scientific image sees persons and also artificial cognitive systems as vast assemblages of postulated units of some kind. In the manifest image, persons are seen as unified centre of representation, deliberation and action, able to reach focused, unified decisions and take focused, unified actions. Much of the paper is devoted to exploring the murkier of the two, the manifest image. It is richer and more diverse than might at first be thought.

## **The Problems Facing Cognitive Science**

The problems currently facing cognitive science have been a subject of much discussion lately. As a number of speakers observed at the workshop on Cognitive Science Education at the meetings of the Cognitive Science Society in 1994, even after forty years of work, cognitive science is still far from having a unified research programme. In this regard, the neurosciences make an interesting comparison. Though much more recent as a self-identified activity, their major international organization has over ten times the members of the Cognitive Science Society and its members aren't constantly worrying about its progress. The problems run very deep: sometimes it is difficult to know how to take or assess the very claims that cognitive science makes.

Doubtless there are many reasons why cognitive science is facing problem but one of them, I think, is this. Using the terms introduced by the

philosopher Wilfred Sellars thirty years ago (Sellars 1963), we have two images of the person. One is the image we have of persons as highly unified centres of representation and intention. This is the image that dominates in everyday life and in philosophy, the social sciences, law, etc. The other is the image of persons as consisting of vast assemblies of postulated tiny units of some kind: neurons, syntactic structures, stimulus-response arcs, or whatever. This is the image that dominates in the sciences of the person, including cognitive psychology and cognitive science. Sellars called the former the *manifest image* and the latter the *scientific image*. One of the reasons for the current difficulties facing cognitive science, it seems to me, is that the scientific image of cognitive science has not as yet made much contact with the manifest image of everyday life and the social sciences.

Why is this a problem, it will be asked. Why can't the two activities be perfectly happy going their separate ways. After all, each has its own proper domain of discourse and a lack of good bridges from one domain of discourse to another is not always a problem, not by any means. It is a problem in the special case of cognitive science, however. We live in the manifest image. It is where we start when we set out to think about ourselves and others. Thus, if we do not know in some measure how to connect claims made within the scientific image to our manifest-image experience of ourselves and others, to that extent we do not know how to take or assess those claims at all.

To understand the kind of connection between the manifest and the scientific image of the person, it might be helpful to have an example. In the last decades, we have developed many marvellous inference and decision systems. Now ask, what do these systems tell us about the activity of me and other people making a decision: the effort to think the situation through and reach a decision, then the effort (the same or a different kind of effort?) to adhere to the decision in the face of some unforeseen complication or -- a different kind of case -- against a strong desire to do otherwise? Very little; the operating principles of inference systems, etc., make almost no contact with decision-making as we find it in persons.

This gap appears in many different places. Cognitive science has had less impact on social policy, interpersonal relationships, rules and structures for regulating interpersonal practice, the law, and many other aspects of social life than might have been expected. Where it has had an impact, it has often not been in the way one might have expected, by illuminating important aspects of individual or social life, but by providing new and powerful prosthetics, in the way word processors have had an impact on writing. (Education is a partial exception, and one could think of others.) Again, one reason, in my view, is that no way has been found to apply the conceptions of human cognition developed within cognitive science to people as we encounter them in everyday life, society, philosophy, the law, etc. More simply, we have no idea how to use these conceptions to understand

ourselves. Note that the problem is not about introspection. How to apply cognitive science's notions of a cognitive system to others is as much a mystery as how to apply it to ourselves.

To forestall a possible misunderstanding, let me hasten to add that I am not setting the stage for any form of mysterianism. (Flanagan (1991) calls philosophers like Nagel and McGinn who urge that consciousness, subjectivity or whatever is somehow beyond the ken of rational modelling New Mysterians.) To the contrary, if there is a problem here, it is a problem to be solved, not a mystery to be venerated. Philosophers, Charles Taylor for example, have taken up some aspects of the issue.

### **The Manifest and Scientific Images of the Person**

We seem, then, to have two broad conceptions of the person and no good way to connect them. As I said, Sellars took a large first step towards getting these two conceptions clearer thirty years ago when he introduced the notions of the manifest image and the scientific image in his very well-known paper, 'Philosophy and the scientific image of man' (1963). To repeat, the manifest image is the image of the human person of ordinary moral, social, and interpersonal life. What makes the scientific image special is that it starts from some postulated tiny unit of analysis and views the person or the human cognitive system as a vast assemblage of these tiny units. The leading theory in the scientific image nowadays is that persons are a vast assemblage of neurons tied together in complex biochemical and informational relationships.

Sellars treats these two images of the person in concert with a similar pair of images of the rest of the natural world but we will consider them in isolation here. He urges that the two images are radically different from one another.

'The' scientific image is in fact more than one image. A number of scientific pictures have had an effect on our conception of persons, including: the neurological picture, computational, information-processing pictures such as the condition-action rule of classical production systems, and more recently the connectionist, distributed-representation picture. If we extend our considerations to include historical ones, the number becomes even larger. In addition to the neural, informational and weighted node units of more recent theorizing, there is the neural hydraulics of Descartes, the neuron and its quantities of energy of Freud's *Project for a Scientific Psychology*, and behaviourism's stimulus-response arc.

These pictures are very different from one another in many ways, of course, but they all have a similar general structure, one that makes them all quite different in much the same way from the manifest image. Thus no harm is done by talk of *'the'* scientific image.

To draw out how the manifest and the scientific images are different a bit more clearly, let us start with the point that in the manifest image, the basic unit of analysis is not some tiny postulated entity. The basic unit of analysis in the manifest image is in fact nothing less than the whole person, a being conceived as able to observe, make decisions, identify itself with things, enter into relationships with others, govern itself by standards, and so on. Much of social science and virtually all of practical, interpersonal or social activity take the whole person as the basic unit of analysis in this way. They start from the person as a unit and focus on what moves these units, how they relate to other such units, how they relate to things in the nonpersonal world, and so on. More generally, the manifest image starts with the person as manifest in everyday life. By contrast, the scientific image takes as its basic building block unobservable entities and unobservable processes of one kind or another, postulated to exist in order to *explain* features of the manifest (Sellars 1963, p. 19). The idea behind this image is that the large basic unit of the manifest image consists of vast numbers of the smaller basic unit of the scientific image.

It is important to understand that these two images are not images of *parts* of the person. As Sellars puts it, they are not two halves of a single picture. Rather, they are,

*two pictures of essentially the same order of complexity, each of which purports to be a complete picture of man-in-the-world. (1963, p. 4).*

### **The Nature of the Manifest Image**

In broad outline at least, the scientific image is fairly straightforward and what we have said already is enough to introduce it. The manifest image is more complicated.

At root, the manifest image of something is simply the image we have of that thing in everyday life. Sellars' picture of the manifest image is more complex than this, however, because he also includes in it everything we can discover about the thing without postulating simpler, unobservable objects to explain it, everything we can discover by observing correlations, etc. Thus with persons, we have been able to discover, for example, that anger is correlated with insults, contentment with good relationships, adult

disturbance with childhood traumata, and many other things by observing and organizing correlations. Moreover, many of these discoveries go well beyond what was contained in our original commonsense conception.

Sellars sometimes equates the manifest with the observable (p. 19). Even if we include the introspectible in the observable as Sellars does (p. 19), this characterization is too restrictive. There is a great deal in our manifest image of persons that is not observable or introspectible, very straight-forward things like character and levels of ability, for example. Neither is observable or introspectible, yet both are clearly aspects of our manifest image of the human person.

So what does characterize our manifest image of the person? One natural suggestion is that the manifest image is the arena of psychological explanation in the language of intentionality. The scientific image would then be characterized as the arena of mechanistic, i.e., non-teleological explanation.

The idea at the heart of psychological explanation is the familiar idea that objects of perception, desire, fantasy, belief, memory, etc. can have meaning for people. As a result, people can think and feel and do things for reasons, not just as the result of mechanistic causes. Reasons in turn have intentionality; they are about something, have an object.

Psychological explanations in terms of reasons are very different from standard mechanistic explanations and it would certainly be true to say that part of the difference between the two images is captured by this difference. But there is more to the manifest image than this. At least one kind of mechanistic explanation also plays a role in the manifest image, namely patterns of correlation as we just saw. They are not postulational and we make constant use of at least of the correlations we have observed in everyday life, which is what makes them manifest-image, yet they are clearly not intentional. The difference between psychological and mechanistic explanation does not exhaust the difference between the manifest and scientific images.

Sellars has a suggestion that may take us further: he says that the manifest image is simply the framework in which we encounter ourselves (p. 6). That is to say, it is the framework within which we experience, reflect on, relate to, and interact with ourselves and one another. This suggestion leaves room for two things that the previous suggestion about psychological versus mechanistic explanation left out. One was the point that beliefs about correlations are also a part of the manifest image. The second is this. In addition to a picture of psychological content and attitudes, which is what the distinction between the two kinds of explanation focuses on, our manifest image of persons also gives us a picture of the thing that has this content, takes up these attitudes. More succinctly, the manifest image is not

just an image of content and attitudes, it also contains an image of the subject of content and attitudes. And it contains an image of an agent, a being that forms intentions, makes decisions, and originates actions.

The subject and agent of the manifest image has a quite specific character. We can capture at least part of what was missed by the suggestion that the manifest image is the realm of psychological or intentional explanation if we can describe it.

Probably the most important feature of this part of our manifest image of the person is the one just sketched:

- (1) Persons are subjects and agents -- centres of representation, imagination, reflection, and desire, originators of intentions, decisions, and actions.

By 'action' we mean not simply behaviour but behaviour that is the result of the formation of an intention and the taking of a decision.

As pictured in the manifest image, to be a subject of perception and representation and an agent of intention and action requires, in turn, that:

- (2) Persons guide themselves by reasons, states and events that work by providing motivation -- motivation to accept or reject beliefs, motivation to feel in one way or another, motivation to act or not to act in certain ways. In the manifest image, we picture persons as beings whose motions are generally caused by motives, not by non-intentional causal forces.

Finally, as pictured in the manifest image,

- (3) Among the most important motivators or reasons for action are emotions -- fears, feelings of affection, feelings of gratitude, resentments, hostilities - and biologically-based desires -- hungers, lusts, feelings of discomfort, and so on.

These three features capture a lot of what is distinctive about persons as we picture them in the manifest image. As thus conceived, when a being has these features, it also has a distinctive constellation of powers, dispositions, and abilities. So the next question is, what are the other things that go with these three features like? It would be impossible in a short space to give anything like an exhaustive list of them but here are a few representative examples.

*1. Making an effort.* Sometimes it is easy to reach a decision, but sometimes it takes effort -- effort to understand a situation, effort to figure out what to do, effort, sometimes, to resist temptation and keep to a decision

(‘I really should reread three more pages of this paper but I would so much like to quit for the night’), and even effort to overcome obstacles, both human and non-human. Moreover, try as hard as we might, sometimes we do not manage to do what we decided to do. What is making an effort like? Are all the exertions of effort just sketched of one kind or a number of different kinds? What is the agent who makes these various efforts like? It is hard to say.

2. *Unity of focus.* Making choices requires something else that is central in our manifest image of the person. I will call it *unity of focus*. To see what it is like, start with the better known unity of consciousness (UC). We can define UC more formally as follows:

The unity of consciousness (UC) =df. (i) a representing in which (ii) a number of representations and/or objects of representation are combined in such a way that to be aware of any of these representations is also to be aware of other representations as connected to it. (For more on UC, see Brook 1994, Ch. 3.)

This is one form of UC. It is a matter of being aware of a whole group of representations at the same time. Another consists of being aware of oneself as the common subject of those representations.

Clearly UC in both forms is an important part of the manifest image of the person. I think a kind of unity found on the volitional side is even more central to what we conceive a person to be, however, what I call unity of focus. We conceive of persons as beings able to focus their intentional resources on courses of action. They can focus on a number of considerations at the same time and weigh up their implications. They can focus on a number of alternative courses of action at the same time and assess them against one another in the light of one's desires, moral beliefs, wishes for other people, etc. They can bring these considerations together to form an intention and choose a course of action. And they can focus their intentional resources on carrying out that course of action, against obstacles, conflicting desires, and so forth. The unity of focus involved in these activities is something more than just unified consciousness of representations.

3. *Anticipating a future as one's own.* I have a striking ability to imagine a future person as myself, to imagine me having the experiences and doing the actions that I suppose he or she will have or do, even when I suppose at the same time that I am connected to that person by all manner of unusual connections and/or lack of connections. Williams explored this phenomenon in a series of interesting thought-experiments in (1973, esp. Ch. 3).

Unified awareness of self and anticipating a future as one's own requires that one refer to oneself. Reference to self is worth a quick sketch

even in a short account because it has some very interesting features.

*4. Reference to Self.* Persons can refer to themselves. Moreover, when I refer to myself, I know that it is to myself that I am referring. The indicators that we use to perform these acts of referring have some unusual semantic properties, properties different from those found in virtually any other indicator, so unusual that Castañeda (1966), for example, does not think that they are true indicators at all. He calls them quasi-indicators. Shoemaker calls the kind of reference in question *self-reference without identification*:

My use of the word 'I' as the subject of [statements such as 'I feel pain' or 'I see a canary'] is not due to my having identified as myself something [otherwise recognized] of which I know, or believe, or wish to say, that the predicate of my statement applies to it. (1968, p. 558)

That is to say, I am aware of myself, and of myself as myself, without inferring this from any other feature of myself. If so, that the referent is myself is something I know independently of knowing anything else. If that is so, in turn, I must be able to refer to myself as myself independently of 'noting any quality' in myself, as Kant put it (1781, A355). If all that is true, finally, the first-person pronouns are semantically quite unusual.

All of (1) to (4) seem to be either parts of our manifest image of the person or natural extensions of it.

Two things are striking about the resulting conception of the person. One is its central importance in practical and especially interpersonal and social life. The other is how little anybody, philosopher, cognitive scientist or anybody else, has managed to say about it. We have made especially little progress with capturing any of (1) to (4) in a postulational, mechanistic model. That is as true of contemporary neuronal, computational, and connectionist models as it was true of the models of earlier times. This gap, to bring us back to where we started, is one reason, in my view, why cognitive science has had as little impact on activities that revolve around the manifest image of the person as it has had.

Much the same gap may underlie the sorry state of functionalism as a high-level model of the mind at the moment. In much the way that we do not know how to map any mechanistic model onto manifest image phenomena, we cannot map the claims of the various flavours of functionalism onto manifest image phenomena.

### **Bridging the Gap: Dennett and Fodor**

To attempt to address this gap or these gaps between manifest phenomena and our various models of the mind, various strategies are unfolding. The 'mind-mapping' work so central to current neuroscience (MRI imaging, etc.) is one. The rapidly-increasing empirical research on consciousness of the past few years is another. Two others are found in the work of two contemporary philosophers, Daniel Dennett and Jerry Fodor. I will restrict my comments to the last two approaches.

Dennett's strategy for addressing the gap between manifest image phenomena and what we have been able to capture in our scientific image models is to try to eliminate the manifest image, at least as anything worth taking seriously. The manifest image is merely that: an image, an interpretation we have constructed because it is very economical way of making some sense of certain complex patterns of behaviour. Thus all the philosophers and others who thought that they were doing metaphysics of the mind have merely been engaged in arcane studies of the self-image we have collectively constructed of ourselves. Oh dear! Unfortunately, I do not think this strategy can get us very far.

There is a feature of the manifest image underlying all the aspects of the manifest image sketched above that I have not yet mentioned. As we picture ourselves, we are each of us a kind of homunculus, a conscious entity who is at the centre of and the unifier of our representational and volitional world. Dennett (1991) has mounted the most thorough attempt to date to undermine this idea, to persuade us that there nothing in people to correspond to this picture. Thus his approach to the gap, over-simplified a bit, is to urge that the problem is our sense that there is a problem. There is no problem. We think there is only because we misinterpret 'phenomenology' in certain natural ways, phenomenology being the way we appear to ourselves and to others.

The problem with Dennett's strategy, in my view, is that it merely shifts the problem. The puzzle now becomes the phenomenology itself: what could produce such a sense as our sense that each of us is an homunculus? And how could we have been led so astray -- how could a conception of such profound, even ineliminable social importance as our manifest conception of the person be at its core nothing more than a huge mistake? (This in outline is all I think most eliminative strategies achieve; they merely shift the problem.)

By contrast, Fodor *is* willing to mechanize the manifest image, or at least to try. His approach is to chip away at the edges: sort out a syntax module here, a vision module there, and postpone the rest to a better day. The problem with this approach is that, while the chips are eminently worth knocking off and we have learned a great deal about the abstract and in the

case of vision even the neural structure of some encapsulated, nonconscious subsystems by doing so, the central person, the big homunculus, has not been touched. It remains as much of a mystery as ever. All Fodor's and others' discoveries about the various nonconscious cognitive subsystems in a person have done is to shrink the range of the unified subject and agent, shrink the homunculi. They have done little or nothing either to discharge it or tell us what it is like.

Why do we need to `reduce' the homunculi? Why not adopt Dennett's strategy and just ignore it or explain it away? For this reason. On the one hand, we cannot dispense with it; we could as soon dispense with the language of love, law, psychology, motive, feeling, and representation. On the other hand, as the Churchlands and others have pointed out, the manifest image does not take us very far. In fact, it works at all only so long as the person being explaining is functioning well. Introduce any amount of cognitive or emotional damage or breakdown and the manifest image instantly `claws the air', as Paul Churchland has put it (1984, p. 46). And so on. Abandoning the manifest image is not one of our options (on this point I do not agree with the Churchlands) but abandoning the drive to capture it in a mechanical model that illuminates and helps us understand it is not an option either.

In the light of these needs, it is striking that most researchers in cognitive science right now are either manifest image extenders or manifest image ignorers. There is very little work on reducing the manifest image going on. Philosophers are prominent in the first group: they spend most of their time trying to find ways to capture manifest image phenomena in extensions and developments of the same kind of phenomena. These folks worry hardly at all about how their ideas might be realized in a computational system or any other kind of mechanical, well-characterized postulational system.

The second group includes most of the rest of cognitive science, the researchers who spend their time developing ever deeper and broader postulational models. Few of these people give much thought to how their models can capture manifest image phenomena. Sometimes one finds both sides in a single researcher, quite unconnected. I am thinking of cognitive psychologists I have known, for example, who work with pure postulational models in their labs, yet delight in tossing around notions such as the unity of consciousness outside it.

What is missing on both sides is an effort *to bring the two images together*, to find postulational, computational accounts that illuminate phenomena of our everyday experience of ourselves. And that, I contend, is one reason why cognitive science is in the state it is in. In the absence of robust roots connecting our scientific image models to our manifest image

experience of ourselves, we not only do not know how to assess these models, we do not even know how to take them. As Dennett once said on a related issue, a meeting of minds would seem to be in order.

## REFERENCES

Brook, A. 1994. *Kant and the Mind*. Cambridge University Press.

Castañeda, H.-N. 1966. "'He': A Study in the Logic of Self-Consciousness". *Ratio* 8, pp. 130-57.

Churchland, P. 1984. *Matter and Consciousness*. Bradford Books/MIT Press

Dennett, D. 1991. *Consciousness Explained*. Little, Brown.

Flanagan, O. 1991. *The Science of the Mind*, 2nd ed. Bradford Books/MIT Press.

Fodor, J. 1983. *The Modularity of Mind*. Bradford Books/MIT Press.

Freud, S. 1895. Project for a scientific psychology. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans and ed J. Strachey. Hogarth Press and Institute of Psycho-Analysis.

Kant, I. 1781. *Critique of Pure Reason*. Trans. Norman Kemp Smith as: *Immanuel Kant's Critique of Pure Reason*. Macmillan, 1963.

Sellars, W. 1963. Philosophy and the scientific image of man. In: *Science, Perception and Reality*, Routledge and Kegan Paul, pp. 1-40.

Shoemaker, S. 1968. Self-reference and self-awareness. *J. of Phil.* 65 no. 20 (1968), pp. 555-567.

Taylor, C. 1985. *Philosophical Papers*, 2 vols. Cambridge University Press.

Williams, B. 1973. *Problems of the Self*. Cambridge University Press, esp. Ch. 3, The self and the future.