J. Expt. Theor. Artif. Intell. 8(1996)383–397

The case against AI from imagistic expertise

SELMER BRINGSJORD

Dept. of Philosophy, Psychology and Cognitive Science, Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY USA

ELIZABETH BRINGSJORD

Dept. of Educational Psychology and Statistics, State University of New York at Albany, Albany, NY and Russell Sage College Troy, NY, USA

1. AI and the fountain of youth

Herein we present a non-technical case for the view that 'robust' imagistic expertise, like its symbolic counterpart,[1] can't be replicated by way of the machinery available to AI, whether of the connectionist or logicist variety. Focusing on imagistic representation and reasoning (R&R) isn't untimely: after years of dealing almost exclusively with rather primitive symbolic forms of knowledge R&R (KR&R), AI has come round to the view, defended in ancient times by Aristotle (1941 trans., Tye 1991) and recently by Kosslyn (1980, 1983, 1994) and others (Chandrasekaran *et al.* 1993, Funt 1980, Lindsay 1989), that KR&R need not, and indeed ought not, be restricted to symbolic, logical form. That AI has traditionally represented the objects of knowledge and belief in symbolic form is wholly uncontroversial; the clearest instance of this approach, and the one which, as a matter of fact, has dominated the field, is to imagine that a robot's knowledge is represented in first-order logic (Charniak and McDermott 1985, Genesereth and Nilsson 1987) or an equivalent (e.g. frame-based systems), and that reasoning consists in deduction over that knowledge. Now, however, AI researchers are facing up to the fact that humans routinely represent and reason about information not only in symbolic form, but also in **imagistic** (or **diagrammatic** or **pictorial**—we harmlessly conflate these terms herein) form, and that this mode of R&R has considerable utility (see Chandrasekaran *et al.* 1993 for a remarkably long list of tasks which effectively tap pictorial R&R (PR&R)). AI's change of heart would seem to augur well for the field: it would seem to imply that not only tomorrow's robots, but today's expert systems, will soon possess and exploit the artificial counterpart to human imagistic R&R.

However, over a decade ago, a number of thinkers confidently claimed that such representation and reasoning would be forever beyond a digital computer. For example, here's what Ned Block said on the matter:

> If the pictorialist view is right, then the human brain deploys representations (and processes operating over them) of a sort not found in digital computers (whose representations are paradigms of descriptional [= symbolic; S.B./E.B.] representations). So digital computers would not be able to process information in the manner of humans (though of course they might nonetheless be able to simulate human information processing). (Block 1981, p. 4)

And, from Hubert and Stuart Dreyfus:

> [C]omputers, programmed as logic machines, cannot use images or any picturelike representations without transforming them into descriptions. (Dreyfus and Dreyfus 1986, p. 90)

If we assume that whatever representation and reasoning AI deploys must be implemented on a digital computer (or, put another way, must be implement*able* on a Turing Machine), then if Block and the Dreyfuses and others are right, and if AI's turn from pure symbolicism to symbolicism-infused-with-PR&R will be sustained, as certainly appears likely,[2] it follows that AI is embarking on a search for something as elusive as the fountain of youth. And, in fact, although the Block/Dreyfus position is now almost universally rejected, it does seem to *us* that Ponce de León and AI are indeed in lockstep. We defend this view with some new argumentation that takes account of work generally regarded to have forever silenced Block/Dreyfus worries. In particular, our argumentation confronts head-on Kosslyn's (1994) seminal attempt to position robust imagistic abilities firmly within a computational conception of the mind and brain.

2. The imagery debate contestants

Let's start with the admission that the Block–Dreyfus reasoning cited above is not invulnerable: it can, in fact, be attacked on at least three fronts. First, it's not at all clear what 'imagistic R&R,' means. In what sense are mental images like pictures?—given that, as everyone must agree, we don't literally find pictures in our heads. Second, absent a careful account of the concept, it's surely not *obvious* that digital computers, or Turing Machines, can't handle imagistic R&R. Third, it's far from self-evident that 'imagism' (or 'pictorialism'), whatever it turns out to be, is *true*.

Each of these avenues of attack has proponents. For example, a number of thinkers have taken the third avenue: they have claimed that pictorialism is false, that the only sort of representations humans use, at bottom, are symbolic, language-like ones (Dennett 1981, Pylyshyn 1981). This position has come to be known as **descriptionalism**. Of course, descriptionalism used as a weapon against the Block/Dreyfus argument might be massively counter-productive, since descriptionalism seems *itself* to imply that AI's turn toward a hybrid symbolicist-pictorialist approach is misguided—for on descriptionalism everything can be done, and perhaps done efficiently, by focusing exclusively on symbolic R&R.

Other thinkers, pleased with AI's move towards hybrid R&R, have combined the first and second avenues of response to Block/Dreyfus worries: they have said that when pictorialism is clarified, it turns out that PR&R *can* be implemented on digital computers. For philosophers, the *locus classicus* of this rebuttal is perhaps Michael Tye's (1988, 1991) position that the computational correlate to our mental images are patterns of charged cells in arrays within digital computers, or something similar. For AI researchers, cognitive scientists, psychologists, and so on, the supreme hybrid view is championed by Kosslyn (1994) who in his recent monograph, *Image and Brain*, affirms and defends the view that human PR&R can be captured by appropriately configured neural nets.

For people closer to the engineering side of AI and Cognitive Science, there is an abundance of work which seems to vindicate the Kosslyn/Tye position by incarnating it in the form of some rather impressive computer programs (Funt 1980, Lindsay 1989). Kosslyn himself, as we shall see, regards such systems to be confirmation of his theory. He also claims to find confirmation in neuroscientific research of a sort we'll examine below.

3. The main argument

Here, without further ado, is the main argument in our Block/Dreyfus-resurrecting case:

A_1

(1) Expert imagistic R&R, which involves the creation, manipulation and contemplation of TEMIs,[3] can't be captured in traditional symbolic representation schemes.

(2) Any form of R&R that can be captured in neural nets can be captured in (e.g.) the state diagram scheme often used to specify Turing Machines.[4]

(3) The state diagram scheme often used to specify Turing Machines is a traditional symbolic representation scheme.[5]

∴ (4) Genuine expert imagistic R&R, which involves the creation, manipulation and contemplation of TEMIs, can't be captured in neural nets.

(5) If expert imagistic R&R can neither be captured in neural nets nor in traditional symbolic representation schemes, it's expertise AI and Cog. Sci. can never replicate, and never fully understand scientifically.

∴ (6) Expert imagistic R&R is expertise AI and Cog. Sci. can never replicate, and never fully understand scientifically.

A_1 is obviously formally valid. That is, (4) follows from (1)–(3), and (6) follows from (4) and (5). This is easily verified by symbolizing A_1 in first-order logic (where one quantifies over forms of R&R). Premiss (5) merely reflects the uncontroversial view that AI and Cog. Sci., fundamentally speaking, are exhausted by two different orientations, the symbolicist and the connectionist. The former centres around traditional symbolic representation schemes (frames, state diagrams, logics, semantic networks, etc.), the latter around neural nets. Proposition (4) is an intermediate conclusion; (6) is the final conclusion. This leaves (1), (2) and (3). But proposition (3) is unexceptionable: Turing Machines (and schemes reducible to TMs; see note (5)) are a traditional symbolic representation scheme. So everything boils down to (1) and (2). But (2) is a mathematical fact one of us (Selmer) has defended elsewhere at length in the context of AI and Cog. Sci. (Bringsjord 1991). The basic idea behind this proposition is a chain which to our surprise sometimes comes as news to even some veteran AIniks, namely: neural nets can be rendered as cellular automata; cellular automata can be rendered as k-tape TMs; k-tape TMs can be rendered as standard, one-tape TMs (and, of course, TMs can be identified with state diagrams of the appropriate sort). This is why we don't have in neural nets a class of system which exceeds TMs (or other equivalent formalisms, e.g. Register Machines) in fundamental information processing power.[6] It is true that analog chaotic neural nets are more powerful than TMS (Siegelmann 1995), but they correspond to symbol-processing devices resembling TMS (Bringsjord 1993).

So everything seems to hinge on (1). What is there to recommend this proposition? A lot, from where we stand. But of course for you, premiss (1) is pretty much impenetrable! After all, you don't know what we mean by 'TEMI'.

4. Temporally extended mental images

'TEMI' is an acronym abbreviating 'Temporally Extended Mental Image,' and such creatures are probably best introduced by way of a little contradistinguishing. (For more, see Bringsjord and Bringsjord 1994b, Bringsjord 1992b.) Accordingly, let's take a quick look at the sort of representations traditionally employed in programs claimed to possess at least some degree of visual imagery capability. In order to take this look, however, we need to cursorily review the concept of a **logical system**. Given our current space constraints, we are forced to leave this concept rather informal. (See Ebbinghaus *et al.* 1984 for a full account of primitive logical systems. One of us (Selmer) has discussed AI at length in the context of not only logical systems, but the continuous systems of connectionist approaches: Bringsjord 1991.) So, informally put, a logical system $\mathscr{L}$ is composed of an **alphabet** A, a **grammar** G which, when applied to A, yields the set of **well-formed formulas** $G(A)$, a set of **rules of inference** for carrying out proofs over $G(A)$, and a formal specification of truth conditions for when the putative binary relation $\models$ is true of a given formula ϕ and a so-called **interpretation** I. Traditional symbolicist AI has for the most part worked within the logical system known as first-order logic, denoted in the canonical notation for logical systems as $\mathscr{L}_I$. For example, the $\mathscr{L}_I$ formula $\forall x \exists y Gyx$ might mean, on the standard interpretation (for arithmetic) that for every natural number n, there is a natural number m such that $m > n$. And the rule of inference known as **universal instantiation** would allow us to move from this formula to $\exists y Gy2$, where 2 is a constant from the relevant alphabet denoting, as expected, the natural number 2, so that this new formula asserts that there is a natural number greater than 2. According to descriptionalism, *all* of the objects of knowledge and belief for a human and a future robot can be represented as formulas in some computable logical system, and all reasoning is deduction over these formulas.

Now we're ready to take a quick look at the sort of representations typically employed in programs claimed to possess at least some degree of visual imagery capability. (Our objective, recall, is to unpack the notion of a TEMI; we're on the route towards doing just that.) The core representational idea in such systems is a rather easily conveyed kernel (Funt 1980, Johnson-Laire 1983, Larkin and Simon 1987, Lindsay 1989): First, consider Figure 1.

ABC is a right triangle; a perpendicular runs from the right angle to the opposite side, giving rise to two triangles *DAB* and *DCA*. Now, establishing the proposition that area of $ABC =$ area of $DAB +$ area of DCA within a logical system is extraordinarily complicated (Tarjan 1971). But, as Lindsay points out.

> [The inference to this conclusion] is direct, immediate and compelling simply from an examination of the diagram .... [This inference] is of course an instance of the whole being equal to the sum of its parts. (Lindsay 1989, p. 240)

If you now imagine that some computer programs, when run on digital computers, allow for diagrams like Figure 1 to be represented, and propositions of this sort to be reached via routes which both capitalize on such representations and are strikingly shorter than ordinary deduction, you have one hand a serviceable concretization of what we call **AI Pictorialism**. And you've also acquired a sense of what we call, in contradistinction to TEMIs, **simple images**, or **simple diagrams**, abbreviated for the

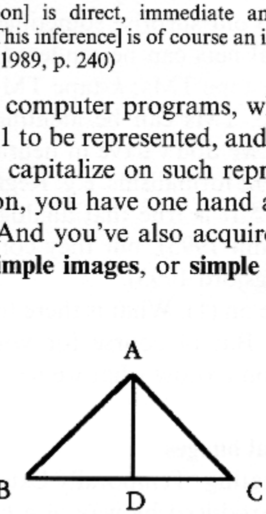

Figure 1.

remaining discussion by **S-D**'s. But can't we be a bit more precise about S-D's? Must we rely completely on an ostensive definition? There is actually a way to capture S-D's with some precision: we say—having sufficient space only to encapsulate this account—that D is an S-D if and only if D and the diagrammatic reasoning thereon can clearly be fully represented in some logical system $\mathscr{L}$. It should be apparent that Figure 1, and the reasoning thereon which we looked at above, can be wholly recast in a logical system. (After all, we know that Euclidean geometry can be entirely derived from the $\mathscr{L}_1$ axioms of Zermelo–Fraenkel set theory.) And we think it's safe to say that S-Ds are what AI has tended to concentrate on since the seminal experiments of Shepard and Metzler (1971) got the ball rolling.[7]

All this is beginning no doubt to look like a massive diversion. What is a TEMI? A TEMI is the sort of image which those who are expert in imagistic fields routinely process. That is, TEMIs are not simply 'visual aids', like drawings [and whether or not S-D's are used as drawings, they do appear to be always *usable* as such; consider, in this connection, Euclid's (1956; trans.) own reasoning]; they are much more robust. For example, consider a screenwriter. Many mature screenwriters are able to write complete drafts of movies 'in their heads'. That is, they are able to watch a movie play out before their mind's eye, from beginning to end, replete with dialogue, props, intonation, scene changes, camera angles, and so on. They then transcribe what they 'see' and 'hear', and the result is a script, which can then be used as a blueprint by producers and directors. The first draft of the American film *Tin Men*, for example, was apparently dictated, start to finish, by the screenwriter and director Barry Levinson.

Such experts have yet to be rigorously studied by cognitive scientists. On the other hand a group of people with related abilities have (supposedly) been studied. We refer to those with the capacity for **eidetic imagery**, those who can, for example, 'examine a picture for several moments and then cast onto a surface an image with positive colour and high detail that lasts several minutes' (Stromeyer and Psotka 1970). These are people, in other words, who can, often at will, generate mental images as clear and distinct as seen real-world scenes, and can then scrutinize the images like you and I might examine a real scene. (Such an ability enables astounding feats of memorization: Luria's (1968) famous mnemonist apparently used eidetic imagery.) So, for example, Stromeyer and Psotka have this to say about one of their amazing subjects:

> [She] can hallucinate at will a beard on a beardless man, leaves on a barren tree, or a page of poetry in a known foreign language which she can copy from bottom line to the top line as fast as her hand can write. (Stromeyer and Psotka 1970, p. 347)

Unfortunately, the eidetiker whose feats are described here by Stromeyer and Psotka was apparently not tested by anyone else. And, more generally, some cognitive psychologists (Gummerman *et al.* 1992) have questioned the very existence of eidetic imagery. Though at least one of us (Selmer) is quite comfortable with the standard claims for eidetic imagery, for the sake of argument we are quite willing to concede that eidetic imagery is chimerical. We cheerfully make this concession because surely there is no denying that many authors (and artists, scientists, etc.) routinely tap a non-preternatural imagistic expertise in order to do what they do. One author who happened to explicitly report about his use of mental imagery was the great dramatist Henrik Ibsen, who said:

> I have to have the character in mind through and through, I must penetrate into the last wrinkle of his soul. I always proceed from the individual; the stage setting, the dramatic ensemble, all that comes naturally and does not cause me any worry, as soon as I am certain of the individual in

> every aspect of his humanity. But I have to have his exterior in mind also, down to the last button, how he stands and walks, how he conducts himself, what his voice sounds like. Then I do not let him go until his fate is fulfilled. (Reported in Fjelde 1965, p. xiv)

In order, then, to get a sense of what TEMIs are, and of what's involved in processing them, we ask you to have in mind a person with Ibsen's imagistic power, someone able to create rich and vivid mental images, *and to stretch such images through time, and to manipulate them in detail so as to produce a robust narrative*.

5. Defending A_1's key premiss (1)

We return now to the issue of what recommends A_1's first premiss. We have four argument-chains for (1); space constraints allow us to present only two here. Each of these four argument-chains corresponds to an observation encapsulable by way of some rough-and-ready prose:

- O1 It's just a brute, undeniable fact that while we can reduce diagrams like Figure 1, and indeed (as we shall indicate below) all the diagrams mastered by AI, to symbolic representations, we can't carry out this reduction for TEMIs. Hence, the burden of proof is on those who maintain that the Block/Dreyfus worry is a carcass.
- O2 TEMIs have a subjective, or qualitative, 'feel' to them, and as such include information which no one knows how to represent in symbolic form (in fact, a number of thinkers have argued directly that subjective qualities of experience utterly defy symbolization: Bringsjord 1992a, 1994b, Jackson 1982, Kripke 1971, Nagel 1974, Searle 1992).[8]
- O3 Unlike the rather rigid status of diagrams like Figure 1, TEMIs are astonishingly elastic. Not only authors, but nearly all of us, can easily picture, say, a Bangkok building in its normal setting which suddenly changes into a rocket ship that blasts off into outerspace. How can this elasticity be captured in an approach which, for example, treats images as diagrams stored as charged cells in an array?[9]
- O4 Though we can't reduce a TEMI to formulas in some logical system, we can of course go a long way toward capturing such images with natural language, which is what authors using TEMIs are in the business of doing. But the natural language involved is itself the richest, most complicated natural language we know of, and the question of whether it can be rendered in a computable logical system is an open one.

We give here the arguments corresponding to O1 and O2. We begin with the argument for O2, which is

A_2

(7) TEMIs include 'what-it's-like-to-____' information.

(8) 'What-it's-like-to-____' information isn't capturable in any symbolic representation scheme.

∴ (9) TEMIs aren't capturable in any symbolic representation scheme.

You may find this argument rather mysterious, for you may not be familiar with the Nagelian (1974) 'what-it's-like-to-be-an-X' locution. A partial explication of this locution, along with a synopsis of the well-known Jackson (1982)–Kripke (1971)–Nagel (1974) argument that deploys it, has recently been provided by John Searle (1992):

> Consider what facts in the world make it the case that you are now in a certain conscious state such as pain. What fact in the world corresponds to your true statement, "I am now in pain"? Naïvely, there seem to be at least two sorts of facts. First and more important, there is the fact that you are now having certain unpleasant conscious sensations, and you are experiencing these sensations from your subjective, first-person point of view. It is these sensations that are constitutive of your present pain. But the pain is also caused by certain underlying neurophysiological processes consisting in large part of patterns of neuron firing in your thalamus and other regions of your brain. Now suppose we tried to reduce the subjective, conscious, first-person sensation of pain to the objective, third-person patterns of neuron firings. Suppose we tried to say the pain is really "nothing but" the patterns of neuron firings. Well, if we tried such an ontological reduction, the essential features of the pain would be left out. No description of the third-person, objective physiological facts would convey the subjective, first-person character of the pain, simply because the first-person features are different from the third-person features. Nagel states this point by contrasting the objectivity of the third-person features with the what-it-is-like features of the subjective states of consciousness. Jackson states the same point by calling attention to the fact that someone who had a complete knowledge of the neurophysiology of a mental phenomenon such as pain would still not know what a pain was if he or she did not know what it felt like. Kripke makes the same point when he says that pains could not be identical with neurophysiological states such as neuron firings in the thalamus and elsewhere, because any such identity would have to be necessary, because both sides of the identity statement are rigid designators, and yet we know that the identity could not be necessary. (117–118)[10]

This conveys the gist of the argument. But it will be helpful to both anchor things with a thought-experiment, and to set out a version of the argument with premises and inferences explicit:

The protagonist in our thought-experiment is a Jacksonian character created elsewhere for our version of the argument in question, which appears in Bringsjord (1992a). The character is Alvin, an AInik who lives and works in an isolated laboratory. Suppose, indeed, that Alvin, for the past five years, has been an absolute recluse, that during this time he has hardly had any contact with other humans, that what contact he has had has all been of a professional, scientific nature, and so on. Alvin, during this time, has mastered the purportedly complete computational specification of human mentation. Suppose, in addition, that he has never encountered a long lost friend—Alvin has never even had an experience remotely like this.

Now, one day Alvin leaves his drab lab and encounters a long lost friend, and thereby learns what it feels like 'on-the-inside' to meet a long lost friend. 'So *this*,' he says to himself, 'is what it feels like to meet a long lost friend in the flesh, to see once again that gleam in her eyes, the light her hair seems to catch and trap....' etc., etc. The corresponding argument in Bringsjord (1992a) is too complicated to reproduce here. So we'll adapt to Alvin's situation a more informal but elegant and powerful statement of the Jacksonian argument given by Jacquette (1994).[11]

A_3

(10) To know everything knowable about a psychological state is to have complete first- and third-person knowledge of it.

(11) Alvin, prior to his first first-person long-lost-friend experience, knows everything knowable about meeting long lost friends from a third-person symbolic perspective.

(12) To know everything knowable about meeting long lost friends from a first-person perspective implies knowing what it's like to meet a long lost friend in the flesh.

(13) Alvin, prior to his first-person long-lost-friend experience, doesn't know what it's like to meet a long lost friend in the flesh.

(14) If what-it's-like-to-____ information is capturable in some symbolic representation scheme, then [if Alvin, prior to his first-person long-lost-friend experience, knows everything knowable about meeting long lost friends from a third-person symbolic perspective, then, prior to his first first-person long-lost-friend experience, he knows everything knowable about meeting long lost friends].

∴ (15) Alvin, prior to his first-person long-lost-friend experience, doesn't know everything knowable about meeting long lost friends. [from (10, (12), (13)]

∴ (16) What-it's-like-to-____ information *isn't* capturable in some symbolic representation scheme. [(from (11), (14), 15)]

Here again the reasoning itself is above reproach: A_3 can easily be symbolized as a proof in the propositional calculus. But please don't misunderstand. Although we consider this argument not only formally valid, but veracious as well, and though we share this opinion with Jackson, Jacquette, Kripke, Searle and others, we recognize that the debate concerning Alvin and his kind is unlikely to stop. In this paper we don't give an *exhaustive* presentation of our case against AI from imagistic expertise. (Such a case would include, among other things, the full defense of A_3 spelled out in large part by one of us (Selmer) elsewhere: Bringsjord 1994b, 1992a, b.) However, premiss (7) is apparently affirmed by Kosslyn (1994, p. 405) himself, who concedes that mental imagery, especially robust mental imagery like TEMIs, has an inherent emotional component. And premiss (8) is certainly the upshot of, if not a proof, then certainly a *formidable* deductive argument (namely A_3). At any rate, there are our (1)-supporting observations which have nothing to do with fanciful creatures like Alvin. We focus, in the remainder of the paper, on O1. Here's how the argument corresponding to this observation runs:

A_4

(17) No symbolization of TEMIs, and the creation, manipulation and contemplation thereof, is in the literature; nor, despite considerable, sustained effort directed at achieving such symbolization, is such symbolization forthcoming—in fact there is no hard evidence to support even the weak position that such symbolization will *eventually* arrive.[12]

(18) If a certain type of object of thought, and an ability to process this type, resists symbolization to the degree that there is not even any hard evidence to support the position that such symbolization will *eventually* arrive, then one ought to affirm the view that such symbolization can't be achieved.

∴ (19) One ought to affirm the view that the symbolization of TEMIs, and the creation, manipulation and contemplation thereof, can't be achieved.

This argument is straightforward; it is also provocative–not because (18) will be controversial (on the contrary, this seems to be an eminently plausible principle, one which could even be put more circumspectively without precluding the core *modus ponens* in question), but rather because most of those thinking about AI and Cog. Sci. in connection with visual imagery believe that work since the Block/Dreyfus objection has buried this objection forever. What work? There is a long list of computer systems generally regarded to mark the symbolization of visual images. We've already mentioned two (Funt 1980, Lindsay 1989), but the complete list is quite a bit longer; it includes: Baron 1985, Gardin and Meltzer (1989), Glasgow *et al.* (1992), Julstrom and Baron (1985) and Kosslyn and Schwartz (1977, 1978). Unfortunately, all of this work, as well as that which is in the works, confirms (17): all of this work includes only the symbolization of S-Ds, that is, only the sorts of 'diagrams' and 'images' which are *obviously* symbolizable. (This claim is something you may well have anticipated, given our assertion above that an 'image' like Figure 1 is essentially the only sort of image AI has managed to represent.) This point can be made by simply going to the images featured in the work itself. As a first example, consider the discussion of mental imagery Dennett (1991) offers us in his *Consciousness Explained.* This is a lengthy discussion, but the key point for our purposes is easily extracted: it is simply that the 'images' Dennett credits robots (e.g., the well-known SHAKEY) and computer programs with creating, manipulating and 'contemplating' are impoverished. They are 'images' which are obviously symbolizable; in short, they're S-Ds. To make the point vivid, let's evoke the sort of images, whether found in robot or human, with which Dennett is concerned.

Consider the 3 by 3 grid shown in Figure 2, and then image it. Now, 'write' the following three words down in the columns of your imaged grid: GAS, OIL, DRY. Finally, attempt to 'read off' the horizontal words formed by this little manoeuvre. Dennett claims that it's hard to 'see' the horizontal words; that's the point he tries to drive home.[13] The point for us is that these simple and static images can obviously be symbolized in such a way as to entirely capture their content. (In the case of Dennett's challenge, the 'images' in question are once again Euclidean and elementary, and hence provably capturable in $\mathscr{L}_I$ systems sufficient to yield at least the bulk of classical mathematics. In this connection, see Ebbinghaus *et al.* 1984).

We imagine some readers saying: 'You conveniently ignore the fact that Dennett's little game is just that: a *little* game. It's a straw man. What about the less 'toy worldish' types of images central to the work in AI and neuroscience which has at least purportedly overthrown the Block/Dreyfus worry?'

Unfortunately, such work never transcends Dennett's scheme. The AI part of the literature in question revolves around 'images' like those shown in Figure 3—cubes (Julstrom and Baron 1985), strings (Gardin and Meltzer 1989), and arrays (Glasgow *et al.* 1992), nothing even remotely like a TEMI, nothing that can't be easily and completely captured in traditional symbolic schemes. None of this work provides the

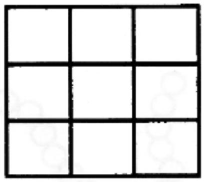

Figure 2.

slightest reason for holding that a TEMI—say that which corresponds to a director's 'internal rehearsal' of a production of *The Tempest*—can be represented in symbolic terms.

Cognoscenti may feel that our TEMI/S-D distinction has an impressive pedigree in Piaget and Inhelder's (1969, 1966) distinction between two broad categories of mental images: **reproductive** and **anticipatory**. Unfortunately, much as we'd like to claim Piaget as a fan, the reality, alas, is that his dichotomy is, even under the most relaxed of readings, at best an approximation of our own. Reproductive images are 'limited to evoking sights that have been perceived previously' (Piaget and Inhelder 1969, p. 71); anticipatory images 'envisage movements or transformations as well as their results, although the subject has not previously observed them (as one can envisage how a geometric figure would look if it were transformed)' (Piaget and Inhelder 1969, p. 71). From this characterization it follows immediately that the mental images sparked in subjects by Shepard and Metzler's (1971) famous geometric figures are, *contra* our earlier classification of them, TEMIs (on the assumption—the only candidate assumption if Piaget's work is to be considered ancestral with respect to ours—that TEMIs are to be identified with anticipatory images). The problem, more generally, is that reproductive images apparently allow for the richness of TEMIs, and anticipatory images include those which are too 'austere' to count as TEMIs (as in the case of Shepard and Metzler's geometric figures). Piaget's distinction considered in a broader context that includes not only our TEMI/S-D distinction, but also AI, is similarly unhelpful (for us), but fascinating. This is so because some of Piaget's work on the reproductive-anticipatory distinction involved experiments in which subjects were faced with challenges arguably met by those AI systems having 'imagistic' capability. For example, Funt's (1980) WHISPER system is capable, at some level, of anticipating the result of moving simple objects—anticipation, if Piaget is right, beyond the grasp of children younger than five.[14] Although, interestingly enough, other aspects of Piaget's work on mental imagery bolster the position we advocate herein,[15] let's return to the main thread of the paper, by asking not how our case might be *helped* (by appeal to the seminal work of a great psychologist), but how it might be *hurt*. What move might those skeptical about our case make?

Well, the most powerful rebuttal those unwilling to affirm (17) can muster would seem to involve arrays. It might be said that if the arrays are sufficiently complex, robust scenes, and manipulations thereof, become possible. Such thinkers will probably draw our attention to real-world scenes which have been digitized, and hence rendered in a form representable as an array. (Of course, this move conflates external versus internal representation, something we overlook for the sake of argument.[16]) Unfortunately, this is all one big bluff we're quite prepared to call. Where, pray tell, is the array-theoretic representation of an 'internally-played' *Tempest* (or anything similar)? Where, in fact, is the array-theoretic representation of anything even

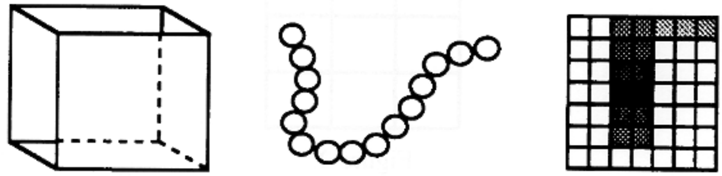

Figure 3.

remotely *like* such a TEMI? Those who would complain that these are mere rhetorical questions and then go on to search feverishly for a shred of evidence in favour of the view that array theory is sufficiently expressive, would do well to keep in mind that array theory is just first-order extensional set theory—and who thinks that a Shakespearean TEMI can be represented in, say, Zermelo–Fraenkel Set Theory?

But what about neuroscience? Do we find here research which dispels the Block/Dreyfus concern? In neuroscience, too, we find but 'images' the symbolization of which is trivial; we find only S-Ds yet again. Consider, for example, the sorts of images at the heart of Kosslyn's (1993, 1994) brain activation research (see Figure 4). Subjects in one of these studies were shown grids with letters represented in them. They were then instructed to create mental images corresponding to these grids. Finally, subjects were shown 'empty' grids, save for one 'X'ed' location; and they were to consult their mental images in order to ascertain if the X fell within the letter (they were told each time what the letter in question was; in Figure 4 the letter is of course 'F'). (PET scanning revealed that imagery activates topographically mapped cortex, but this isn't germane to the issues at hand for us.) It's obvious that this study revolves

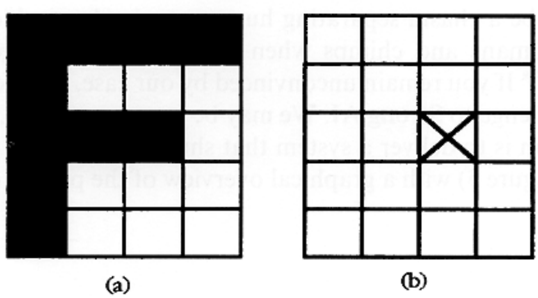

Figure 4. Subjects studied a grid-with-letter, such as (a); were told to image what they had seen, and were then, after being presented with grids like (b), asked whether the 'x'ed' box fell within the letter.

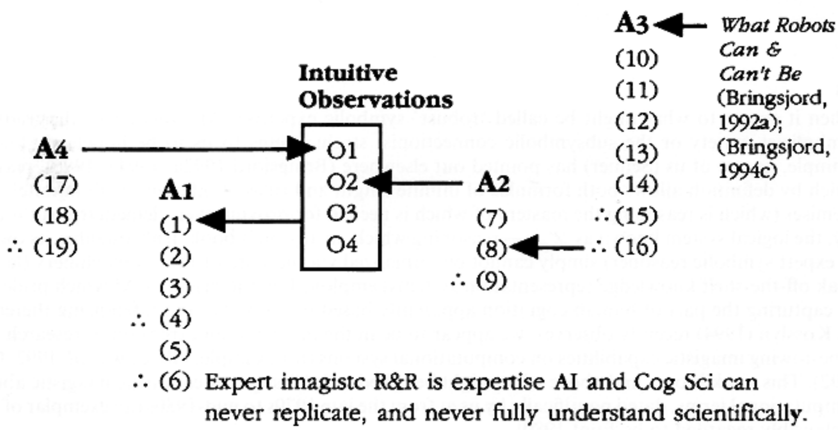

Figure 5. Architecture of our case against AI from imagistic expertise. Note that in the present paper the arguments corresponding to O3 and O4 are not presented.

around images and diagrams which are S-Ds. And this study isn't idiosyncratic. Quite the opposite, really. In fact, it's tempting to say that neuroscientists and neuropsychologists have a fetish not only for S-Ds, but specifically for matrices and grids (see, for example, Attneave and Curlee 1983, Cornoldi *et al.* 1991, Kerr 1987). In sum, neuroscience deals with images which, like all their relatives in the AI/Cog. Sci. literature, pale beside TEMIs: they are images the symbolization of which is transparent. So proposition (17) seems to us to more than hold its own.

6. Concluding remarks

It's important to read our overall position, (6), with care. Doing so reveals that our case against AI is really a case against *Strong* AI. We're not saying that robots who capitalize on S-Ds to get their work done will never be more than creatures of fiction. Quite the contrary: we think AIs will arrive who write not belletristic fiction, but good stories, good stories the generation of which *doesn't require the creation, manipulation and contemplation of TEMIs* (see the computer systems designed and implemented by S. Bringsjord and Ferrucci (forthcoming).[17] But we *are* pessimistic about the prospects for the arrival of a robotic Henrik Ibsen. Put another way, in the sphere of imagery there may forever be a chasm separating humans and robots which is analogous to that separating humans and chimps when it comes to language (Bringsjord and Bringsjord 1994a).[18] If you remain unconvinced by our case, then feel free to view our argument as a challenge to Strong AI. We may be wrong, and if so, the way to deliver a decisive refutation is to deliver a system that shuts us up.

We conclude (Figure 5) with a graphical overview of the part of our case presented above.

Acknowledgements

We're indebted to Seng-Beng Ho (whose unpublished 'The Processing and Representational Nature of Mental Imagery' is a masterful comprehensive analysis of mental imagery in a computational context); David Boles, Dr R. Sadananda, Mark Bickhard, and a number of anonymous referees. And of course we're indebted to Robert Campbell, whose efforts made this issue of *JETAI* possible.

Notes

1. When it comes to what might be called 'robust' symbolic expertise, AI, whether of the traditional symbolicist variety or the subsymbolic connectionist strain, seems to us to be hopelessly lost. For example, as one of us (Selmer) has pointed out elsewhere (Bringsjord 1992a, 1994a, 1995), reasoning which by definition allows both formulas of infinite length and rules of inference with infinitely many premises (which is reasoning the mastery of which is needed for proving even elementary truths about, say, the logical system known as $\mathscr{L}_{\omega_1\omega}$—reasoning which, in anyone's book, truly qualifies its owner as an expert symbolic reasoner) simply cannot be formalized via the state-of-the-art machinery (let alone weak off-the-shelf knowledge representation systems) employed by the brand of AI which prides itself on capturing the part of human cognition apparently based on formulas and inferencing thereon.
2. As Kosslyn (1994) recently observes, we appear to be in the middle of an explosion of research aimed at bestowing imagistic capabilities on computational systems (for example, Glasgow *et al.* 1992, Omori 1992). This work is over and above what might be called AI's 'first go' at capturing imagistic ability in computational terms, dated pontifically by us at from the late 1970s to mid-1980s (an exemplar of which is plausibly regarded to be Funt 1980).
3. Short for 'temporally extended mental images,' a label explained momentarily.
4. An excellent hands-on introduction to the state diagram take on Turing Machines can be obtained via Barwise and Etchemendy's (1993) *Turing's World*, interactive software that comes with a manual

serving as a brief introduction to computability theory. B&E follow Boolos and Jeffrey (1980), a more detailed study. See http://www.rpi.edu/~ brings for a repository of Turing's machines on the web.

5. Of course, as such representation schemes go, Turing Machines are rather 'low-level'. 'High-level' representation schemes would include semantic networks, first-order logic, frames, and so on—all of which could be recast as TMs. At any rate, it's undeniable that Turing Machines (and even, for that matter, simpler automata—such as finite state automata) are a traditional symbolic representation scheme.
6. Although connectionism and logicism are equivalent mathematically, and though as a result we find the clash between these camps to be a red herring (Bringsjord 1991), we agree with thinkers who find connectionist systems inadequate in the face of higher-level symbolic reasoning. Connectionist systems and techniques, we predict, will prove suitable for lower-level cognition—locomotion, perception, and the like, while logicist systems and techniques will prove suitable for at least some aspects of higher-level cognition, such as that required in mathematics. (Of course, by our lights, neither camp will ever deal adequately with any number of challenges, such as the one—capturing the 'meaning' of a TEMI—described in this paper.)
7. Alert readers will detect here a presupposition on our part, namely that our internal representation of Shepard and Metzler's (1971) stimuli are S-Ds. We haven't the space to defend this presupposition, but we urge those disinclined to affirm it to look again at the geometric shapes at the heart of S&M's experiments. These shapes are all constituted by simple line segments drawn in Euclidean 3-space.
8. Of course, qualia aren't restricted to TEMIs.
9. Kosslyn (1994) calls this property of visual images 'malleability.' We don't specify and defend O3 herein, so readers of his book will have to judge for themselves, without our 'help', whether or not he shows that this property can be possessed by neural nets, and hence possessed as well by symbolic formalisms.
10. Of course, to the extent that this argument is sound—and we think it is—it applies to subjective experience of *any* kind, not just that associated with TEMIs.
11. Jacquette's (1994) argument here, like the one presented in (Bringsjord, 1994c), is designed specifically to surmount Dennett's (1991) reply to Jackson's (1982) version.
12. We readily concede that there is *some* evidence for the arrival of such symbolization. For example, the fact that some S-Ds can be symbolized might fairly be read by some as constituting such evidence. But this evidence is hardly *hard* evidence. Consider, for illumination of the point, the following analogy. Suppose that the issue is whether first-order logic is sufficiently expressive to formalize a set S of mathematical statements. Certainly the fact that first-order logic is up to the task of capturing some mathematical statement $p \notin S$ provides some evidence that FOL is sufficiently expressive to formalize S. But no one would regard the evidence here to be *hard*. (After all, if this is hard evidence, then there is hard evidence that every mathematical statement can be formalized in first-order logic!).
13. As a matter of fact, we think Dennett takes this example much too seriously. With a little practice, most of my (Selmer's) students managed to solve these grids rather quickly. We suspect that those with good visualization abilities could become *strikingly* adept at solving them. (Certainly those blessed with eidetic imagistic powers could do as well with visualized grids as mere mortals do with inscriptive ones, but there is no need for us to invoke eidetikers here: Selmer's students make the point just fine.)
14. Children younger than five, if presented with the problem of anticipating a slight movement of one square placed on top of another, cannot produce a drawing corresponding to what will happen if the top square is moved. (Piaget and Inhelder 1969, p. 74).
15. For example, Piaget held that mental imagery included the internal sensations which in today's terminology would be regarded qualia (Piaget and Inhelder 1969, p. 69), and he gave a number of arguments for the Bringsjordian view that mental imagery is not just a fancy version or prolongation or perception (see Section IV. in Piaget and Inhelder 1969).
16. A fuller and more aggressive version of the rebuttal under consideration would insist that adding a temporal dimension to static images is no problem for AI, and adding detail to the point of even the 'eidetic' is also unproblematic. The problems with this reply over and above the two we mention in the main text (one, the reply conflates external representations with internal ones; two, the reply, given the current state of the art, is just one big bluff) include an unwillingness to come to grips with the subjective experience at the heart of TEMIs, and the fact that even external representation schemes in AI are stretched to the breaking point when used to try to capture information as rich as that seen in TEMIs.
17. These stories will also fail to be accompanied, in the machine, by any of the usual states of mind enjoyed by human authors.
18. We're not begging any questions here in favour of a Chomskyan view of language. For the proposed analogy to be enlightening, we need only assume something which even those who ascribe *bona fide* linguistic capacity to chimps must affirm, namely that no amount of training is going to allow a chimp to interview John Updike. Likewise, it may be that no amount of time and effort will eventuate in a computational system blessed with TEMIs.

References

Aristotle (1941) *On Memory and Recollection*, 45a 5, 450a 7, and 451a; 19. In R. McKeon (ed.) *The Basic Works of Aristotle* (New York: Random House).

Attneave, F. and Curlee, T. E. (1983) Locational representation in imagery: a moving spot task. *Journal of Experimental Psychology: Human Perception and Performance*, **9**: 20–30.

Baron, R. J. (1985) Visual memories and mental images. *International Journal of Man-Machine Studies*, **23**: 275–311.

Barwise, J. and Etchemendy, J. (1993) *Turing's World 3.0: An Introduction to Computability Theory* (Stanford, CA: CSLI Publications).

Block, N. (1981) *Imagery* (Cambridge, MA: MIT Press).

Boolos, G. S. and Jeffrey, R. C. (1980) *Computability and Logic* (Cambridge: Cambridge University Press).

Bringsjord, S. (1991) Is the connectionist-logicist clash one of AI's wonderful red herrings? *Journal of Experimental and Theoretical Artificial Intelligence*, **3**(4): 319–349.

Bringsjord, S. (1992a) *What Robots Can and Can't Be* (Dordrecht: Kluwer).

Bringsjord, S. (1992b) CINEWRITE: an algorithm-sketch for writing novels cinematically, and two mysteries therein. *Instructional Science*, **21**: 155–168.

Bringsjord, S. (1993) Toward non-algorithmic AI. In K. T. Ryan and R. F. E. Sutcliffe (eds) *AI and Cog Sci '92*, in the Workshop in Computing Series (New York: Springer-Verlag) pp. 277–288.

Bringsjord, S. (1994a) Computation, among other things, is beneath us. *Minds and Machines*, **4**(4): 469–488.

Bringsjord, S. (1994b) Searle on the brink, *Psyche*, **1**(5) [32 paragraphs]. URL:ftp://hcrl.open.ac.uk/pub/psyche/volume1/psyche-94-1-5-propertydualism-1-bringsjord.txt.

Bringsjord, S. (1996) An argument for the uncomputability of infinitary mathematical expertise. In K. Ford, R. Hoffman and P. Feltovich (eds) *Expertise in Context* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Bringsjord, S. and Bringsjord, E. (1994a) Animal communication of private states doesn't illuminate the human case. *Behavioral and Brain Sciences*, **16**(4): 645–646.

Bringsjord, E. and Bringsjord, S. (1994b) Can AI accommodate imagistic expertise? In *Proceedings of the Second International Conference on Expert Systems for Development* (Los Alamitos, CA: IEEE Press), pp. 36–41.

Bringsjord, S. and Ferrucci, D. (forthcoming) *Artificial Intelligence, Story Generation, and Literary Creativity* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Bringsjord, S. and Zenzen, M. (forthcoming) *In Defense of Uncomputable Cognition* (Dordrecht: Kluwer).

Chandrasekaran, B., Hari Narayanan, N. and Iwasaki, Y. (1993) Reasoning with diagrammatic representations: a report on the Spring Symposium. *AI Magazine*, Summer.

Charniak, E. and McDermott, D. (1985) *Introduction to Artificial Intelligence* (Reading, MA: Addison-Wesley).

Cornoldi, D., Cortesi, A. and Preti, D. (1991) Individual differences in the capacity limitations of visuospatial short-term memory: research on sighted and totally congenitally blind people. *Memory and Cognition*, **19**: 459–468.

Dennett, D. (1981) The nature of images and the introspective trap. In N. Block (ed.) *Imagery* (Cambridge, MA: MIT Press).

Dennett, D. (1991) *Consciousness Explained* (Boston, MA: Little, Brown).

Dreyfus, H. L. and Dreyfus, S. E. (1986) *Mind Over Machine* (New York: Free Press).

Ebbinghaus, H. D., Flum, J. and Thomas, W. (1984) *Mathematical Logic* (New York: Springer-Verlag).

Euclid (1956) *The Thirteen Books of Euclid's Elements*, trans. Thomas Heath (New York: Dover).

Fjelde, R. (1965) Foreword in Ibsen, H. (1965) *Four Major Plays* (New York: New American Library).

Funt, B. V. (1980) Problem solving with diagrammatic representations. *Artificial Intelligence*, **13**: 201–230.

Gardin, F. and Meltzer B. (1989) Analogical representations of naive physics. *Artificial Intelligence*, **38**: 139–159.

Genesereth, M. and Nilsson, N. (1987) *Logical Foundations of Artificial Intelligence* (Los Altos, CA: Morgan Kaufmann).

Glasgow, J. I., Fortier, S. and Allen, F. H. (1992) Molecular scene analysis: crystal structure determination through imagery. In L. Hunter (ed.) *Artificial Intelligence and Molecular Biology* (Cambridge, MA: MIT Press) pp. 433–458.

Gummerman, K., Gray, C. and Wilson, J. M. (1992) An attempt to assess eidetic imagery objectively. *Psychonomic Science*, **28**(2): 115–118.

Jackson, F. (1982) Epiphenomenal qualia. *Philosophical Quarterly*, **32**: 127–136.

Jacquette, D. (1994) *Philosophy of Mind* (Englewood Cliffs, NJ: Prentice Hall).

Johnson-Laird, P. N. (1983) *Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness* (Cambridge, MA: Harvard University Press).

Julstrom, B. A. and Baron, R. J. (1985) A model of mental imagery. *International Journal of Man-Machine Studies*, **23**: 313–334.

Kerr, N. H. (1987) Location representation in imagery: the third dimension. *Memory and Cognition*, **15**: 521–530.

Kosslyn, S. 1980) *Image and Mind* (Cambridge, MA: Hard University Press).
Kosslyn, S. (1983) *Ghosts in the Mind's Machine* (New York: Norton).
Kosslyn, S. (1994) *Image and Brain* (Cambridge, MA: MIT Press).
Kosslyn, S., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L. and Buonanno, F. S. (1993) Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, **5**(3): 263–287.
Kosslyn, S. and Shwartz, S. P. (1977) A simulation of visual imagery. *Cognitive Science*, **1**: 265–295.
Kosslyn, S. and Shwartz, S. P. (1978) Visual images as spatial representations in active memory. In E. M. Riseman and A. R. Hanson (eds) *Computer Vision Systems* (New York: Academic Press).
Kripke, S. (1971) Naming and necessity. In D. Davidson and G. Harman (eds) *Semantics of Natural Language* (Dordrecht: Reidel) pp. 253–355, 763–769.
Larkin, J. and Simon, H. A. (1987) Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, **10**: 65–100.
Lindsay, R. K. (1988) Images and inference. *Cognition*, **23**: 229–249.
Luria, A. R. (1968) *The Mind of a Mnemonist*, trans. by Lyn Solotaroff (New York: Basic Books).
Mendelson, E. (1987) *Introduction to Mathematical Logic* (New York: Wadsworth & Brooks).
Nagel, T. (1974) What is it like to be a bat? *Philosophical Review*, **LXXXIII**: 435–450.
Omori, T. (1992) Dual representation of image recognition process: interaction of neural network and symbolic processing. In *Proceedings of the International Symposium on Neural Information Processing*, pp. 50–53.
Piaget, J. and Inhelder, B. (1966) *L'Image Mentale Chez L'enfant* (Paris: Presses Universitaires de France).
Piaget, J. and Inhelder, B. (1969) *The Psychology of the Child* (New York: Basic Books).
Pylyshyn, Z. (1981) Imagery and Artificial intelligence. In N. Block (ed.) *Readings in Philosophy of Psychology*, Vol. 2 (Cambridge, MA: Harvard University Press).
Searle, J. (1992) *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).
Shepard, F. N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science*, **171**: 701–703.
Siegelmann, H. T. (1995) Computation beyond the turing limit. *Science*, **268**: 545–548.
Stromeyer, C. F. and Psotka, J. (1970) The detailed texture of eidetic images. *Nature*, **225**: 346–349.
Tarjan, R. E. (1971) *An efficient planarity algorithm.* Report STAN-CS-244-71, Stanford, CA, Stanford University.
Tye, M. (1988) The picture theory of mental images. *The Philosophical Review*, **XCVII**(4): 497–520.
Tye, M. (1991) *The Imagery Debate* (Cambridge, MA: MIT Press).
Willard, D. (forthcoming) Self-verifying axiom systems and the incompleteness theory.