Phenomenology and the Cognitive Sciences **1:** 389–394, 2002.

Representational parts

RICK GRUSH[1] AND PETE MANDIK[2]
[1]Department of Philosophy, University of California, San Diego, La Jolla, CA 92093-0119, USA (E-mail: rick@mind.ucsd.edu); [2]Department of Philosophy, Washington University, USA (E-mail: Pete@twinearth.wustl.edu)

In this reply we claim that, contra Dreyfus, the kinds of skillful performances Dreyfus discusses *are* representational. We explain this proposal, and then defend it against an objection to the effect that the representational notion we invoke is a weak one countenancing only some global state of an organism as a representation. According to this objection, such a representation is not a robust, projectible property of an organism, and hence will gain no explanatory leverage in cognitive scientific explanations. We argue on conceptual and empirical grounds that the representations we have identified are not weak unprojectible global states of organisms, but instead genuinely explanatory representational parts of persons.

Growing dissatisfaction with traditional artificial intelligence over the past decade or so has resulted in an increasing number of research projects that take up the cause of anti-representationalism. We think this is wrong-headed: not because we are enamored with traditional AI, but rather because the problem with traditional AI was not its representationalism, but its vision of representations as logical symbol structures, and its conviction that intelligence is best understood as problem solving via manipulation of such symbol structures. Accordingly, we think Dreyfus' anti-representational challenge is misguided. There are a number of points at which one could quarrel with his article. For example, even if one grants phenomenology the explanatory authority Dreyfus grants it, it is clear that his description is incomplete. Often when playing chess, one not only visually recognizes certain configurations and opportunities on the board, and perhaps feels the right thing to do. But there are often times that one imagines making non-actual moves: *What would happen if I moved the black bishop there? Then my opponent would likely do that, and then I could do that.* Such bouts of reasoning are undoubtedly common not only in playing chess, but in many cases of problem solving, and are curiously absent from Dreyfus' phenomenology. What one is exactly *not* doing in such a case is letting the world be its own representation; rather, one is letting an internal representation (image, whatever) stand for the real chess board.

Granted, in such a case, the representer will be using his learned skills to recognize opportunities on the imagined board configuration. But this in no way undercuts the fact that what the skills are engaged on in such cases are representations. We are not forced to try out our ideas in the actual board. If we lacked the ability to represent the board, if we had to let the world be its own representation, we would be so forced.

But pursuing such a line is not our immediate concern. Rather, we will try to make a stronger point, that the fact that some activity is a context dependent skill does not entail that that activity is not also representational. To assume otherwise without argument is to commit a common fallacy, as when one argues that the purpose of stop signs is not to get cars to stop, because there is evidence that their purpose is to reduce the number of car collisions. Both can be legitimate purposes. In the present case, the unargued, unrecognized, and false assumption is that an organism's possession of certain embodied skills cannot be that in virtue of which it grasps certain contents. Or to put it another way, that the implementation base of a skill cannot be the vehicle carrying a certain content, i.e. be a representation. Only with this assumption in place do we have the needed dichotomy: skill or representation.

Gareth Evans (1985), for example, argued persuasively that an organism's sensorimotor skills are that in virtue of which an organism is able to grasp spatial content, to represent space. Evans' idea was that for an organism with the appropriate suite of skills, an experience of a certain sort simply presented itself *as* the sort of thing on which it could bring those skills to bear. This shares similarities to Gibson's notion of an affordance, but Evans does not unnecessarily drain the notion of its representational content, as Gibson does.

This should not seem unattractive even to Dreyfus, and indeed is hiding just beneath certain verbal obfuscations like "the world presents itself as . . .", which he exploits shamelessly. Once we recognize that one can think about non-actual chess-board configurations, it becomes clear that what it presenting itself cannot be the world *per se*, but is rather a representation of it. All motivation for preferring "the world presents itself as" over "the world is represented as" dissipates.

It should be clear that we are not completely parting company with Dreyfus on this issue. Even on our suggestion, what is crucial is an organism's mastery of certain embodied skills, and not its possession or manipulation of logical formulae. And the content grasped on the basis of these skills is one that can be specified only by mentioning these skills: they present the world, or part of it, as something on which that or that skill can be brought to bear. For a person with a minimal set of skills, a chunk of wood is seen as graspable,

indeed, *represented* as graspable. Supplement the minimal set of skills with those constitutive of knowing how to play chess, and the same chunk of wood may be seen as, say, a queen, and thus, as capturable. And, we shall want to say, the queen is thus *represented* as capturable.

With these points in mind, perhaps Dreyfus would not be repulsed by our proposal. Admittedly, the proposal described thus far is skeletal and hanging further flesh on it requires resources not available in this short commentary. Our task now is not to articulate the proposal further but instead to defend the proposal against a certain kind of objection.

The objection runs as follows: even if in fact skills are that in virtue of which a creature entertains contents, it does not follow that there are representations. The implementation base of a skill can be massively superposed, and in such a case it would not make sense to talk of a representation at all. In fact, connectionist accounts offer a plausible story for how distinct abilities might nonetheless have superposed physical bases. In such cases, it might be claimed that though there are represent*ings*, there are no represent*ations*. Are we such massively distributed/superposed systems? Probably not. Consider the following.

It seems obvious that many representations used by humans as well as those used by other sophisticated nervous systems have contents whose components are *independently targetable*. What we mean by "independently targetable" is best conveyed by an example. You can imagine the Eiffel Tower as being black, then as blue, then as teal. You can hold the color constant while imagining it now as being very heavy, now as being very light (perhaps you imagine picking it up with one hand, and then imagine grasping it and being unable to budge it). There is evidence that children reach a stage of development where they become much more able to independently vary the represented properties of objects (see Karmiloff-Smith 1992 for a review of some of this evidence). The skillful chess player can counterfactually reason about keeping all but one chess piece in place. While keeping all else constant, the knight is imagined elsewhere. Likewise for the other pieces.

Now there are two ways that such independent targeting can occur. The simple hypothesis is that in such systems, the independently targetable represented properties are carried by physical states which are independently causally targetable. In such a case, one can imagine a rod of the same thickness but different lengths by causally targeting only those physical states carrying length contents, but not causally targeting those carrying thickness contents.

But if the contents are fully superposed, then this will not be possible. For instance in the standard example of superposed representation, the hologram,

such independent targetability is not possible, because one cannot go into the physical hologram itself and alter only one of the represented properties while keeping the others constant. You can't change the position of the bird in the branches of the tree without destroying the representation of the tree.

Consider as an example the following superposed representation. The 3×3 grid of numbers contains a superposed representation of 2 words, "big" and "die". This is extracted by adding the numbers in the appropriate columns or rows,

0	2	2	**d**
2	7	0	**i**
0	0	5	**e**
big			

and taking the letter with that ordinal position in the alphabetic ordering. Thus the first row, with the numbers 0, 2 and 2 adds to 4, and the forth letter is "d".

Now suppose that I want to change the word represented in the final column from "die" to "can", but to do so while keeping the word in the bottom row constant. This is no easy task (go ahead and try). And there are many choices for which it is not possible (one cannot change "die" to "car", for example, and get the numbers to add up right while keeping "big" along the bottom).

The lesson to be drawn from the matrix example is how bloody difficult it is to have both fully superposed representations and independent targetability. In fact, the situation is much worse than this example indicates, because in the matrix, you as the person operating on the matrix, had different representations of the words you were trying to work with. If we imagine a single system that is in its entirety like the matrix, and needs to alter its own structure so as to keep certain aspects of it invariant, we run into a conceptual difficulty – how does the system maintain consistent track of what things are to be held constant, so as to solve for the invariances? In order to maintain those representations constant throughout the change, it will have to have already solved the problem (figured out a way to keep its knowledge of what is to be kept constant throughout the problem-solving episode) before it begins to try to solve it!

But leaving this objection aside, and assuming that somehow the system can pull it off, the questions are how? and why? How could a system do the sort of reconfigurational acrobatics required to alter the entire physical state in just the way required to keep all the contents but one constant? The case of

the 3 × 3 matrix, with only 9 elements and only 2 very simple superposed contents, was difficult enough. Try imagining the task with several billion neurons, and a large number of complex contents (including the contents of memories!! These should not be altered every time I think of anything new). Those who tout fully superposed representations are suspiciously silent about this. The second question is why? Why waste time and effort rewriting all the books in the library every time a new letter arrives in the mail? The obvious ploy will be to avert to some sort of dynamic processes, and some notion of an attractor. But such systems typically take some time to settle into stable states. Furthermore, there are in fact well-known syndromes involving focal lesions to the cortex knocking out specific classes of object representations – so the facts seem to be on the side of the non-superposed camp anyway.

To take stock, the assumption that different components of the content are carried by different physical states is a hypothesis that makes sense, for which there is a great deal of evidence, and whose only competitor seems to face some fairly serious computational and conceptual difficulties. Now we are not saying that in order for a system to be representational at all, it must employ vehicles that are not fully superposed. Rather, our argument is that (i) even if the representations are fully superposed, they are still *representations*, and (ii) it is not likely that the representations of any minimally sophisticated cognitive systems are fully distributed.

In summary, we have tried to make three points: First, much skillful performance requires counterfactual reasoning, where one is not letting the world be its own representation. Second, one can very well embrace the idea that much of intelligent behavior involves the mastery and exercise of skills without becoming an anti-representationalist, by pointing out that skills can be that in virtue of which a content is grasped. Third and finally, we have defended the claim that skillful performance can be representational against the objection that such representations may just as well be massively distributed represent*ings* : things not worth calling representa*tions* after all. We wish to concede that the alleged representings *may* be massively distributed. But this seems to be an overly involved way of parceling out cognitive responsibility that faces some serious challenges.

References

Grush, R. (ed). 1998. Skill and spatial content. *Electronic Journal of Analytic Philosophy* 6. *Special Issue on the Philosophy of Gareth Evans.* http://ejap.louisiana.edu/EJAP/1998/contents.html.

Grush, R. and Mandik, P. (forthcoming). Anti-anti-representationalism. (Draft available).
Karmiloff-Smith, A. 1992. *Beyond Modularity: A Developmental Perspective On Cognitive Science*. Cambridge, MA: MIT Press.
Mandik, P. 1999. Qualia, space, and control. *Philosophical Psychology* 12 (1): 47–51.
Smolensky, P. 1991. Connectionism, constituency and the language of thought. In: B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics*. Cambridge, MA: Basil Blackwell.