Encoding Psychological Knowledge*

Ian Pratt

Abstract

A widely accepted philosophical view has it that our grasp of the *psychological* concepts—*believing*, *desiring*, *intending*, *doubting*, *hoping*, *fearing* and so on—is to be understood as our mastery of a *theory* governing their use—specifically, a theory consisting of that body of commonsense knowledge which David Lewis calls *folk psychology*. In this paper, I aim to show why, and in what sense, it is wrong to understand our grasp of psychological concepts—or our mastery of psychological vocabulary—as knowledge, however implicit, of any theory. My strategy will be to sketch a rival view of the cognitive mechanisms underlying our grasp of psychological concepts, and to show that the operation of those mechanisms cannot effectively be duplicated by our possessing any collection of beliefs, that is to say, by our knowing any theory. In doing so, I will pay particular attention to our knowledge of the typical environmental causes and behavioural effects of our psychological states, and I shall claim that such knowledge arises through processes of observation and simulation.[1]

1 Introduction

The purpose of this paper is to challenge a widely accepted philosophical view concerning the nature of the *psychological concepts*. By "psychological concepts", I mean, centrally, the concepts of *believing*, *desiring*, *intending*, *doubting*, *hoping*, *fearing* and so on. According to the view I want to challenge, our grasp of these concepts is to be understood as our mastery of a *theory* governing their use—specifically, a theory consisting of that body of commonsense knowledge which David Lewis calls *folk psychology*. As Lewis puts it:

*Appeared in: "Encoding Psychological Knowledge", in Clark, A. and Millican, P. (eds.) *Connectionism, Concepts and Folk Psychology: The legacy of Alan Turing*, Oxford: Clarendon Press, 1996. pp. 249–264.

[1] In the time this paper has been in press, my views have developed somewhat. See, for example, I Pratt: "Analysis and the Attitudes" in Wagner, S. and Warner, S. (eds.): *Naturalism: a critical appraisal*, Notre Dame, IN: University of Notre Dame Press, 1993. However, the differences between that article and the present one are primarily matters of detail.

> Think of common-sense psychology as a term-introducing scientific theory, though one invented long before there was any such institution as professional science. Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. ... Add also all the platitudes to the effect that one mental state falls under another—'toothache is a kind of pain', and the like. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us—everyone knows them, everyone knows that everyone else knows them, and so on. For the meanings of our words are common knowledge, and I am going to claim that the names of mental states derive their meaning from these platitudes.[2]

Thus: psychological terms are to be defined in terms of a general theory of common-sense psychology governing their use; psychological concepts are to be analysed by outlining the general form of that theory; and our grasp of those concepts is to be understood as (implicit) knowledge of that theory.

In the sequel, I aim to show why, and in what sense, it is wrong to understand our grasp of psychological concepts—or our mastery of psychological vocabulary—as knowledge, however implicit, of any theory. My strategy will be to sketch a rival view of the cognitive mechanisms underlying our grasp of psychological concepts, and to show that the operation of those mechanisms cannot effectively be duplicated by our possessing any collection of beliefs, that is to say, by our knowing any theory. In doing so, I will pay particular attention to our knowledge of the typical environmental causes and behavioural effects of our psychological states, and I shall claim that such knowledge arises through processes of observation and simulation. The thrust of the paper will be to establish the fundamental differences between the account of our grasp of the psychological concepts as knowledge of folk-psychology, and the rival account I present.

2 A theory of psychological inference

The purpose of this section is to present an account of psychological inference—an account, that is, of how we reason with the concepts of *believing*, *desiring*, *intending*, *doubting*, *hoping*, *fearing* and so on. In subsequent sections, I shall argue that this account undermines the view which sees our grasp of psychological concepts as possession of a theory of folk psychology.

Perhaps a good place to start is with our knowledge of our own psychological states. Sometimes, it seems, we consciously monitor our thoughts through a process of introspection: not only do we believe P, Q, R and desire X, Y, Z, but we also believe that *we believe that* P, Q, R, and believe that *we desire that* X, Y, Z. But not

[2]David Lewis: "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy*, 50 (1972), p.256.

all monitoring of our beliefs takes this conscious and deliberate form. Sometimes, it seems, we *un*consciously monitor the course of our thinking for later recall. For the fact that I can remember much of what believed or desired yesterday (whether or not I still believe and desire these things) suggests that I monitored the occurrence of these thoughts at the time, even though I made no conscious effort to do so. So the first question we shall ask is: how might such monitoring processes be implemented? Or, more precisely: how might beliefs about one's own psychological states—beliefs of the form "I believe P", "I believed P yesterday", "I desire X", "I desired X yesterday" *etc.*, be represented in our brains in order that that these monitoring processes might be made to work?

We can simplify the answer by indulging in a little fictional psychology. Suppose that for you to believe that P is a matter of the tokening, in a suitable part of your brain—let us call it your *belief-box*—of a certain (physical) type of data-structure Δ_P. We shall say that, for you, Δ_P *encodes* the proposition P. Similarly, let us suppose that for you to desire that X is a matter of the tokening, in a suitable part of your brain—let us call it your *desire-box*—of a certain (physical) type of data-structure Δ_X, where Δ_X encodes, for you, the proposition X. And so on for all propositional belief-contents and all psychological modalities. Now we can suggest a way for you to encode propositions about your own current psychological states. To encode "I believe P", take the data-structure Δ_P encoding the proposition P, and *tag* it with some special symbol denoting the modality of belief. By 'tagging', I mean literally altering the physical properties of Δ_P is some small way, so that it is easy to recover, if necessary, the original from the tagged version. To encode the propositions "I believe Q", "I believe R", just tag the corresponding data-structures Δ_Q, Δ_R, with the same symbol denoting belief. Similarly, to encode the proposition "I desire X", take the data structure Δ_X, encoding the proposition X, and tag it with some special symbol denoting the modality of desire. To encode the propositions "I desire Y", "I desire Z", just tag the corresponding data-structures Δ_Y, Δ_Z with the same symbol denoting desire. Similarly for the other modalities: intending, doubting, fearing etc. Let us denote above tagged data-structures by Δ_P^B, Δ_Q^B, Δ_R^B, Δ_X^D, Δ_Y^D, Δ_Z^D *etc*: believing "I believe P" is then a matter of your having a token of Δ_P^B in your belief-box; believing "I desire X" a matter of your having a token of Δ_X^D in your belief-box; and so on.

Given this rather simple-minded representation scheme, it is easy to see how the sort of monitoring mentioned above might work. The presence of the belief P, in the guise of the tokening of Δ_P in your belief-box, can give rise, via the tagging process just mentioned, to the presence of the belief "I believe P" in the guise of the tokening of Δ_P^B in your belief box. Similarly with beliefs about your other psychological states. And there is no reason why the tags thus attached to your mental data-structures cannot be time-stamped, and the resulting data-structures retained for an indefinite period in your belief-box. In this way, a tagged data-structure with a given time-stamp lying around in your belief-box encodes a proposition about one of your *previous* psychological states. Clearly, for you to be able to remember a previous belief P which you no longer hold, we must presume that the tag on the time-stamped data-structure Δ_P^B

protects it from the ravages of whatever processes of belief-revision swept Δ_P away. And indeed, there seems no reason in principle why tagging data-structures should not thus alter the way in which belief-revision processes can latch onto them.

Now if you can monitor when you were in this or that psychological state—say, the belief that P—then you can, in principle, notice the conditions under which those psychological states arose in the past. Whence you can induce generalizations about the typical aetiology of those psychological states. Since the mechanisms described above enable you to form beliefs about the correctness of your earlier beliefs, conclusions about the causes of the belief P, will not, in the main, take the form: "I believe P when and only when it is the case that P", but will rather contain information about the conditions under which you believed P even though (as you now know) it was false, or failed to believe P even though (as you now know) it was true. And there seems to be no limit, in principle, to the sophistication and detail of the knowledge you can garner in this way. Indeed, there is no reason, if the foregoing account of this sort of knowledge is correct, why your knowledge the typical causes of the belief P need make no reference to the proposition P at all: after all, most of my beliefs about hexagons were caused by books, or so I seem to remember.

I have spoken of the ability to form inductive generalizations concerning the causes of a *particular* belief P. But what of generalizations concerning the typical causes of, for example, beliefs *about cats*, or beliefs *about hexagons*, or beliefs *about sub-atomic particles* ? In judging, if you do so judge, that most of your beliefs about hexagons were caused by reading books, you made a judgement concerning all beliefs whose contents belong to a certain class of propositions; yet all you could have observed by means of the introspective processes described above are, in the first instance, beliefs with *particular* contents: the belief "Hexagons have six sides", "The internal angles of a hexagon sum to 4π" and so on. How then might you induce generalizations about the causes of *classes* of beliefs identified by quantifying over belief-contents? Again, we see that that the encoding scheme described above causes no particular problems. For since, by hypothesis, it is easy to recover the original data-structure Δ_P from the tagged version Δ_P^B, those inferential mechanisms by means of which you can determine that a given proposition—e.g. "The internal angles of a hexagon sum to 4π"—is a proposition to do with *hexagons*, can also be deployed to determine that the proposition "I believe that the internal angles of a hexagon sum to 4π" is a proposition about a belief whose propositional content is to do with hexagons. That is: the fact that you can systematically transform Δ_P^B into the untagged data-structure Δ_P means that encoding propositions about your psychological states using such tagged data-structures renders the propositional contents of those beliefs available to the full range of your inferential machinery. In particular, the encoding scheme described above causes no problems in regard to determining logical relationships between the propositional contents of psychological states thus represented.

Two final points concerning your knowledge of your own psychological states. First, just as the tagging and untagging mechanisms postulated above give you access to arbitrarily sophisticated and detailed information concerning the environmental *causes* of your psychological states, those same mechanisms can yield information of similar quality about the behavioural *effects* of your psychological states. You might observe, for example, that an unqualified intention to lift a small object, or to cross the road, or to utter a sentence, standardly issues in your actually performing the corresponding actions; you might observe, for example, that an intention to meet a paper deadline or solve a mathematical problem standardly does not. Second, although I have so far focused on knowledge of the *environmental* causes and *behavioural* effects of psychological states, there is no reason why your introspective mechanisms should be limited to such extra-cranial conditions. There is no reason, that is, why you should not observe that whenever you have such-and-such (types of) beliefs and such-and-such (types of) intentions, then you invariably form such-and-such (types) of desires. Here, the conjunctions you notice are not between a thought and some external condition, but rather, between one constellation of thoughts and another.

So much then for encoding knowledge of your own psychological states. But how do you encode propositions about the psychological states of *others*? And how will that encoding allow you to acquire and reason with such information? The critical insight in understanding our knowledge of other people's minds, I believe, is stated by Hobbes thus:

> But there is another saying not of late understood, by which they might learn truly to read one another, if they would take the pains; and that is, *Nosce teipsum*, *Read thy self*... [which is meant] to teach us, that for the similitude of the thoughts, and Passions of one man, to the thoughts, and Passions of another, whosoever looketh into himself, and considereth what he doth, when he does *think*, *opine*, *reason*, *hope*, *feare* &c, and upon what grounds; he shall thereby read and know, what are the thoughts, and Passions of all other men, upon the like occasions.[3]

That is, we assume that, when it comes to psychological states, the situation with other people will be much as it is with ourselves. If I have found, for example, that chancing on a cat strolling across my path in broad daylight usually causes me to believe that there is a cat in front of me, then I can assume that similar beliefs will be so caused in others. If I observe that my intention to stroke such an animal, or to attract it's attention, usually results in my making such-and-such a sound or holding out my hand in such and such a way, then I might assume similar intentions to underly such actions performed by others.

Nor is it hard to devise a natural method for encoding propositions about the psychological states of other persons. Just as the proposition "I believe that P" is encoded by a tagged version Δ_P^B of the data-structure Δ_P, so let the proposition "S believes that P" be encoded by a differently tagged version ${}^S\Delta_P^B$ of the same data-structure, where this time, the tag incorporates a pointer to your concept of the person S in

[3] Thomas Hobbes: *Leviathan*, (ed. C.B. Macpherson) Harmondsworth, Middlesex: Penguin Books (1968), p.82.

question. The point about such an encoding is that it facilitates the process of one's generalization about the causes (and effects) of one's own beliefs to the case of other people. For, under the suggested encoding, that process is essentially one of systematically transforming data-structures of the form Δ_P^B to suitable data-structures of the form ${}^S\Delta_P^B$. In short: it is a process of tag re-writing. Similar remarks apply to the encoding of propositions about other psychological states—desires, intentions, hopes, fears *etc.*—of other persons. Notice that, as with propositions about your own psychological states, the suggested encoding scheme makes the propositional contents of the represented psychological states readily available to the full range of your inferential mechanisms. So there is no problem, on that encoding scheme, about detecting logical relations between the propositional contents of others' psychological states.

I have so far been speaking of your ability to *observe* your own psychological states, as if your psychological knowledge relied on the constant vigilance of your mind's eye, always on the lookout (or should I say, *look-in*) for evidence for or against sundry generalizations concerning the interactions of your psychological states. But surely, it will be objected, to reason effectively about the thoughts of others (which will in general be different from your own thoughts), you will need to know about the interactions entered into by constellations of thoughts which you yourself have never had, and which you yourself could therefore not have observed by introspection. Thus it might be thought impossible that any amount of introspective vigilance could give you information about the likely evolution of the thoughts of others, and hence impossible that the mechanisms suggested above could give you the capacity to reason about the mental life of anyone whose thoughts are, in the relevant respects, different from your own.

Suppose, then, you have inferred that S believes P, Q, R and desires X, Y, Z, and that you wish now to know how that psychological state will develop. (I ignore other psychological modalities for brevity). Here is one possibility, for some years now a familiar and hotly-debated topic in the philosophy of mind. What if you *imagine* that P, Q, R were true and that X, Y, Z were desirable? For, in so doing, you would *simulate* the mind of one who really does have these thoughts. Extending our fictional psychological model we might suppose that such an act of imagination involves setting up a 'pretend' belief-box containing the data-structures Δ_P, Δ_Q, Δ_R and a 'pretend' desire-box containing the data-structures Δ_X, Δ_Y, Δ_Z. (Recall that a tokening of Δ_P in your regular belief-box would amount to your believing P, and so on.) In addition, you must copy into these pretend belief- and desire-boxes the data-structures corresponding to those of your background beliefs and desires which you assume to be shared by S. Having cloned a suitable set of pretend beliefs and desires, you then let those data-structures interact with each other and with your standard inferential machinery to yield new data-structures (corresponding to new 'pretend' beliefs and desires) which you can then observe using your usual introspective powers as detailed above. Being data-structures whose occurrence in your normal belief- and desire-boxes would constitute your believing P, Q, R,... and desiring X, Y, Z,..., the behaviour of these data-structures will provide a reliable guide to what you would have thought if you had believed and desired those things. Therefore, if Hobbes is right, they will be a reliable guide to what any person *will* think who *does* believe and desire those things. Hence, the 'thoughts' you detect in your simulated cognitive predicament you may then presume to be the thoughts that are likely to come to S.

Thus the ability to imagine different situations, to simulate other people's presumed or putative cognitive predicaments, turns mere observation into experiment. For, in order to know how a given set of beliefs and desires will interact, you need never have observed the effects of such constellations of psychological states in yourself: the fact that you can imagine what it is like to be one who has these thoughts, that you can construct a model of such a person's psychological states, gives you access to all the data you require.

How does our suggested encoding scheme fare when it comes to such simulations? Again, very well. For suppose you encode the above propositions about S's beliefs and desires using the tagged data-structures ${}^S\Delta^B_P$, ${}^S\Delta^B_Q$, ${}^S\Delta^B_R$ (S's beliefs) and ${}^S\Delta^D_X$, ${}^S\Delta^D_Y$, ${}^S\Delta^D_Z$ (S's desires), as suggested. Then the task of constructing a model of S's beliefs and desires is merely one of inserting, into your auxiliary belief- and desire-boxes, the untagged versions of those data-structures: namely Δ_P, Δ_Q, Δ_R and Δ_X, Δ_Y, Δ_Z. And, as we have said, systematically untagging data-structures is, by hypothesis, a simple operation. Similar remarks of course apply to encoding, as conclusions about S's likely thoughts, the results of the simulation: that is just a matter of tagging interesting data-structures produced in the simulacrum with labels indicating that they are the likely future beliefs of S, and of inserting the tagged data-structures into your own belief box.

Much has been written for and against the suggestion that psychological reasoning proceeds by simulation.[4] Yet it is clear that simulation, assuming it does have a part to play in psychological inference, can at best provide only a partial account of how such inferences proceed. For surely, much psychological inference is directed towards such questions as why someone acts in a certain way, what misconceptions he might be labouring under, whether and why he might be lying, what he must have known and how he could have come to know it. And it is hard to see how simulation, which provides no direct means to reason from effects to causes, can play anything more than an ancillary rôle in tackling such problems.[5] But my aim here is not to produce a workable piece of cognitive psychology. Rather, I merely want to illustrate what I believe to be the fundamental simplicity of the problem of encoding psychological knowledge in such a way that a wide range of psychological inferences become possible, at least in principle. For the existence of such encoding schemes has important consequences for the view according to which our grasp of the psychological concepts is to be seen as possession of a theory of commonsense psychology, as I shall now argue.

[4] See for example, R. Gordon: "Folk psychology as simulation", *Mind and Language* 1 (1986), and S. Stich and S. Nichols: "Folk psychology: Simulation or tacit theory?" *Mind and Language*, 7 (1992).

[5] For a brief discussion of these issues, see I. Pratt, I. Leudar and L. Xu: "Understanding detective stories", *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, 1992, pp. 1046–51.

3 The relationship between thought and object: computational hermeneutics

The striking feature of the mechanisms for psychological inference sketched in the previous section is their lack of dependence on any specific features of the data-structures on which they are used. For it seems likely that the tagging and untagging operations which form the core of these mechanisms would work happily on an essentially unlimited variety of data-structures. Think of a computer implementation of these operations: tagging a data-structure might involve, perhaps, pairing it with some symbol in a list; and just about any data-structure can be paired with a symbol in a list. So, if you augment your stock of concepts with a new type of data-structure which enters into all kinds of new interactions with the data-structures already in your head, your psychological inference mechanisms will cope effortlessly with your newly expanded conceptual repertoire. Whatever new propositions P this new data-structure enables you to encode, no sooner do you acquire the capability to entertain the proposition P (by tokening instances of some new data-structure in your brain), than you also acquire the capability to entertain the propositions "S believes that P", "S desires that P", and so on. Moreover, your ability to construct make-believe cognitive predicaments enables you to experiment with the behaviour of your new beliefs *vis-à-vis* collections of your other psychological states as and when required; so there is no need to wait for instances of these new beliefs fortuitously to present themselves before you acquire an extensive ability to reason about them. If your conceptual repertoire expands (by whatever processes people's conceptual repertoires expand) your mechanisms for psychological inference are well-equipped to cope with their owner's new thoughts.

I explained early in the previous section how the suggested method for encoding propositions about what you *previously* believed enabled you to reach conclusions about the extra-cranial causes and effects of these beliefs. In particular, it enabled you to reach conclusions about the causal relationships, if any, between states of affairs in which P is the case and states of affairs in which you *believe* P. Thus, for example, you might have concluded that the belief "This cat is black" is usually caused by the cat in question's being black, unless of course it is dark, or unless the cat is well-camouflaged, or whatever. And, as we observed above, it is because you can *revise* your beliefs that the monitoring of those beliefs (and of their revision) can yield non-trivial conclusions about the relationship between those beliefs and the states of affairs they are about. Clearly also, the particular conclusions you reach concerning this relationship will depend on the way in which you are disposed to revise the beliefs in question. That is, it is *because* your beliefs about the colour of cats get revised in the way they do—*because*, that is, they interact with your perceptions and other of your psychological states in just the way they do—that you come to just the beliefs you do about the typical causal relationship between your belief "This cat is black" and the colour of the relevant cat. So if, as a result of the enlargement of your conceptual repertoire, new types of beliefs arise, the generalizations you formulate about the conditions under which these new beliefs arise—and in particular, about the relationships between whatever states of affairs these new beliefs are about and the beliefs themselves—will depend on the way these new beliefs function, on the way they interact with your perceptions and with other of your psychological states. And, as we have already seen, the processes by which you arrive at these generalizations will work more or less whatever these new beliefs are, and however they function within your brain.

The following radical suggestion then presents itself. Perhaps we should allow more or less *any* new type of data-structure, having more or less *any* new patterns of interaction with your existing beliefs to count as a *bona fide* psychological state: as a belief, desire, intention, or what have you (depending on which 'box' it is in). Perhaps we should say, that is, that any data-structure that can exist in your belief-box counts as a *bona fide* belief, and, *ipso facto*, as a *bona fide* belief *about something*. After all, it would appear that, on the encoding scheme suggested above, any new type of data-structure that can operate in our brains, happily interacting with other data-structures in sensible ways, can unproblematically be made the object of our psychological concepts. That is: if that new data-structure enables us to engage in certain new modes of thought, we can unproblematically think about our engaging in these modes of thought or about other persons' engaging in these modes of thought. So why not take *whatever* new data-structures can be incorporated into our regular cognitive systems to count as regular psychological states?

Suppose, then, that a new type of data-structure Δ is tokened in your belief box. The question seems to arise: what will the belief you thereby have be *about*? That is: what proposition will Δ encode? The suggestion I wish to make—and here we see the radical nature of my proposal—is that, in general, the *only* way for someone to know what proposition a given type of data-structure in your head encodes is for him to *give himself* something like that data-structure, and to instantiate tokens thereof in his own brain. Then of course, he will know what the belief is about, because he will be able to have it for himself. Moreover, once he can have this belief, once, that is, his conceptual repertoire has been suitably expanded, he should not experience any difficulty when it comes to expatiating on the connections between these new beliefs and what they are about. He can just apply his introspective and simulative mechanisms as outlined above.[6]

Looking at things in this way, the possible connections between between beliefs and what those beliefs are about appear to be virtually open-ended, and appear moreover to be so without compromising your mastery of the notion of belief. To put the

[6] Here, complications intrude which, for the sake of clarity, I have suppressed in this paper. It is not correct to say that for me to know what proposition tokens of the physically-typed data-structure Δ encode for you is a matter of my somehow getting tokens of Δ into my own brain. What I should say is that it is a matter of my getting into my brain tokens of some type of data-structure Δ', whose *functional rôle* in my brain is the same as the functional rôle of Δ in yours. But such complications are not my main concern in this paper. I leave it to the reader to make corresponding adjustments throughout, where necessary.

matter in somewhat Augustinian terms: if I ask you about the relationship between a *particular* state of affairs P and the state of affairs in which you believe that P, then, assuming of course that you can understand my question, you can tell me; but when I ask you what the relationship is *in general* between states of affairs and beliefs about those states of affairs, you cannot say. You cannot say, of course, because there is just nothing, at that level of generality, to be said.

To be sure, we may wish to impose certain minimal restrictions on the kind of behaviour which a mental data-structure can be allowed to get away with and still qualify as a belief or other psychological state. For example, we might insist that any new data-structure combine with existing data-structures by means of the standard logical operations, that it participate in all modes of thought—beliefs *and* desires *and* intentions etc.—that it not vandalize one's psychological state by obliterating or scrambling data-structures with which it comes into contact, and so on. But these minimal restrictions imposed by the basic framework of psychological modalities, may still leave vast latitude for novel kinds of cognitive function in general, yielding novel relationships between thought and object in particular.

The view we have arrived at might perhaps be called *computational hermeneutics*. According to that view, *any* data-structure in one's head obeying certain minimal restrictions counts *ipso facto* as a psychological state *about something*. And there is no general way of saying what a given data-structure encodes other than to have that data-structure in one's own head[7]. But there is nothing mysterious about this view—at least, there shouldn't be. For I have been at pains to stress in this paper the fundamental simplicity of the computational mechanisms needed to manipulate the psychological concepts, when the relationship between thoughts and their objects is understood thus.

Let us return to the question of our grasp of the psychological concepts. I have argued that there is nothing in general to say about the relationship between the thought and its object, between the belief that P and P. It follows that that relationship cannot be captured by any set of principles, excepting, of course, an infinite listing containing on the order of one entry for each possible thought-content P. Therefore, there is no psychological 'theory', neither folk nor esoteric, knowledge of whose principles is equivalent to possession of the mechanisms for psychological inference outlined in the previous section. It is in this sense, then, and for foregoing reasons, that it is wrong to take our grasp of the psychological concepts to be a matter of our knowledge of folk-psychology. [8]

[7] See previous footnote.

[8] One important aspect of our grasp of psychological concepts which I have ignored in this paper concerns the allowed variation in how psychological states with the same content can function in different individuals and on different occasions. That some such variation exists is undeniable. Just because I have solved (or failed to solve) a mathematical problem, that does not mean that you will reach the same result, even if you start with the same set of premises. Just because I am willing to generalize from a few data-points or accept a certain explanation as overwhelmingly plausible, that does not mean that you will be so epistemically bold, even if you have the same evidence at your disposal. Peacocke (Christopher Peacocke: *A Study of Concepts*, Cambridge, MA: MIT Press, 1992, pp. 168–171) uses this observation as an argument against a simulation-based account of the nature of the concept of belief.

4 Encoding psychological knowledge

On page 100 of *The Intentional Stance*, Dennett writes:

> An interesting idea lurking in Stich's view is that when we interpret others we do so not so much by *theorizing* about them as by *using ourselves as analog computers* that produce a result. Wanting to know more about your frame of mind, I somehow put myself in it, or as close to being in it as I can muster, and see what I thereupon think (want, do...). There is much that is puzzling about such an idea. How can it work without being a kind of theorizing in the end? For the state I put myself in is not belief but make-believe belief. If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what "comes to me" in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different? In both cases, knowledge of the imitated object is needed to drive the make-believe "simulation", and the knowledge must be organized into something rather like a theory.

As the recent flurry of interest in psychological simulation suggests, however, there is something more to the idea than Dennett here gives it credit for. Simulation does indeed loom large in the picture of psychological inference sketched in section 2; but what is crucial about that picture is not the idea of simulation, but rather, of *observation*. It is because the cognitive mechanisms underpinning psychological inference are fundamentally mechanisms of *observation* that, as we argued in section 3, those mechanisms are not equivalent to the possession of any psychological theory. Simulation, by contrast, is of secondary importance. What simulation does is to turn observation into experiment. It transports us, as it were, to any aspect of those data structures we want to see. By allowing us to observe the behaviour of data-structures in mental situations in which we would not ordinarily encounter them, it removes the need to extract explicit generalizations about the behaviour of our psychological states in unexperienced situations.

Thus, Dennett's analogy, though understandable, proves misleading. Think of simulation and you think of models of bridges or weather systems or nuclear war or whatever: things which work because they have the programmer's knowledge of the relevant principles embodied in the computer code. But psychological inference—at least, the interesting parts of it—is not a matter of applying principles: it is a matter of making observations. The analogy with our knowledge of bridges, not that it is a particularly illuminating one, should run more as follows. If you want to know about all the different bridges in the world—how many struts they incorporate, what colour the

What, exactly, these allowable variations are, how they are recognized, and how they are taken into account in psychological inference, are intriguing problems about whose solutions I could only speculate here. Since such speculations would not affect the overall thrust of this paper, I shall refrain from them.

pillars are, what company has its name stamped on the rivets—then you must go to the bridges and take a look. Since a bridge is, in general, a too big an object to be seen in detail all at once, those details are best seen by walking around and looking at the different parts of the bridge. But there is nothing general and true that can be said about the peculiarities of each and every possible bridge: there is no general *theory* which tells you about all the peculiarities of each and every possible bridge. Just so, if you want to know about the types of thoughts in your mind, then your cognitive mechanisms must tour those thoughts and inspect them one by one. Since the behaviour of thoughts cannot all be seen at once, you have in general, and if possible, to transport yourself, via the process of imagination, to different cognitive situations in which different aspects of your thoughts' behaviour can be observed. But there is nothing general and true that can be said about all the features of the behaviour of each and every possible thought: in particular, there is no general *theory* specifying the relationships between thoughts and their objects, between the belief that P and P.

But, surely, there are some general principles which apply to bridges; might the same then not be true of psychological states? Why, yes! I do not deny that there cannot be any general psychological principles; I only deny that knowing these principles could suffice on its own to endow us with the capacity to engage in psychological inference.

5 Conclusion

I began this paper by sketching a theory of psychological inference. True, the theory was greatly simplified, ignoring many important phenomena and finessing many difficult problems. But my aim was neither to develop a serious and detailed psychological proposal, nor to put forward an "analysis" or "definition" of psychological states or psychological concepts. Rather, it was to make a philosophical point about the way we should understand the relationship between thought and object, in the light of a plausible view concerning our grasp of psychological concepts. The point is this. It transpires that remarkably simple mechanisms, centred on the operations of tagging and untagging of data-structures, can form the basis of a wide range—perhaps the full range—of the functions required for psychological inference. It further transpires that these same mechanisms can gracefully accommodate an open-ended variety of data-structures, constituting, I suggested, beliefs and other psychological states standing in an open-ended variety of relationships to the states of affairs those psychological states are about. That is: the mechanisms underlying our mastery of psychological concepts impose only minimal restrictions on the relationships between thoughts and their objects. Thus our grasp of the psychological concepts—believing, desiring, intending, doubting, hoping, fearing and so on—has the following property. However our conceptual repertoire may expand, whatever new kinds of thoughts we can entertain, we will always retain the ability to embed those concepts and thoughts within the scope of the psychological attitudes, and to gain knowledge of the causal (and other) relationships between those attitudes and what those attitudes are about. Yet we can do this even though there need be next to nothing that can be said in general about this relationship between thought and object. And if the relationship between thought and object really is open-ended in the way I suggest, then to encapsulate that relationship for every proposition we can entertain would require an infinite listing of cases containing on the order of one entry per possible thought-content. Therefore, our mechanisms for psychological inference—those mechanisms, that is, by which we can know, amongst other things, the relationship between thought and object in any case—cannot amount to knowledge of any set of principles (other than an infinite and unprincipled listing)—cannot, that is, amount to possession of any psychological theory.