From Folk Psychology to Naive Psychology

ANDY CLARK
University of Sussex
Brighton, England

The notion of folk-psychology as a primitive speculative theory of the mental is called into question. There is cause to believe that folk-psychology has more in common with naïve physics than with early speculative physical theorising. The distinction between these is elaborated. The conclusion drawn is that common-sense ascription of psychological content, though not a suitable finishing point for cognitive science, should still provide a more reliable source of data than some contemporary theorists are willing to admit.

I. INTRODUCTION

Opening the gate, the postman spots a large, angry-looking Alsatian. He quickly shuts the gate and moves on to the next house. Asked to explain the postman’s behaviour, we might well say “He saw the dog and believed it was going to bite him. Since he didn’t want to be injured, he shut the gate and went on.” In giving such an explanation we would be indulging in what has become known as folk-psychology. For we would be using a certain theory of the mental (viz. one which appeals to combinations of beliefs and desires to explain actions) to make sense of other people’s behaviour. The question before cognitive science concerns the status of this “primitive theory.” For it is very tempting to use our common-sense ideas as a starting point for serious scientific enquiry. It seems quite possible, however, that if folk psychology is a “primitive scientific theory,” it may be a deeply misguided one. Undue reliance on such a theory would thus be a constant source of spurious problems and not a guide to good psychological theorising. This is the position adopted by Churchland (1979), (1981), (1984) and, most influentially, by Stich (1983). Folk-psychology, they fear, is a theory on a par with, say, early cosmology; a product of the campfire speculations of “ancient shepherds and camel drivers” (Stich, 1983, p. 229). Folk-psychology is distinguished from other early and misguided theories only by its dogged persistence throughout the ages. This persistence is said to be all
the more remarkable when we reflect that the story of folk-psychology has been a sad tale of "retreat, infertility and decadence" (Churchland, 1981, p. 74). Perhaps, then, it is high time for cognitive science to throw off its folk-psychological shackles and ring in the changes.

And perhaps not. It is my purpose, in what follows, to present some considerations which prompt a subtly different view of the status of (so-called) folk-psychology. Whilst accepting that folk-psychology constitutes a kind of primitive theory of the mental, I shall attempt to deny that that theory need be thought of as mere folk speculation. Instead, I suggest that we should accord our common-sense ideas of the mental something of the status we accord our common-sense conceptions of the physical. I thus propose that we talk not of folk-psychology but of naive psychology where this is conceived along the lines of Hayes' (1979) notion of a naive physics. The general argument in favour of this conjecture will be that just as a roughly accurate grasp of some basic physical principles is vital to a mobile organism, so too will some roughly accurate grasp of basic psychological principles be vital to a social organism. It is because the competence underlying this basic grasp gets underwritten by evolutionary necessity that I do not share any extreme pessimism about the use of ordinary psychological concepts in cognitive science. "Folk astronomy" as Stich tells us "was false astronomy and not just in detail" (Stich, 1983, p. 229). But naive physics is not, in general, false physics (though it may be basic, incomplete and needs-oriented). Nor, I suspect, is naive psychology in general false psychology (though it may be parochial and radically incomplete). The structure of my discussion will be as follows. I begin (Section II) by describing the thinking behind the characterisation of our ordinary ideas of the mental as a folk-theory. I go on (Section III) to examine the standard reasons for doubting the acceptability of that theory. In Section IV I develop an alternative account which explains the unusual ("stagnant," "parochial") nature of "folk"-psychological thought without reducing it to mere folk speculation. Section V addresses some possible objections to the arguments offered. Some experimental evidence is then considered (Section VI) and a brief summary given (Section VII).

II. THE VOICE OF THE PEOPLE

What is meant by the claim that our ordinary ideas about the mental (e.g., the explanation of the postman's behaviour) involve some kind of theory? One thing that this could be taken to mean is that there exists some well articulated set of laws and principles, together with some paradigmatic cases of their successful application and a programme of further research, such that we could pick out a guiding "theory of the mental" in the same way as we can pick out a theory of optics or the kinetic theory of gases. This, obviously, is not the sense of "theory" which Stich and Churchland have in mind.
Rather, what they have in mind is something which may be better described as an *implicit theory*. A theory may be said to be implicit if the laws and principles which constitute it have to be inferred from the *use* of the putative theory in generating predictions and explanations. This, as we shall see, is precisely the way that Churchland invites us to conceive of the theoretical status of folk-psychology. We should note at once that not all folk-theories are *implicit* theories in this sense. The theories of alchemy and early cosmology with which we are invited to compare folk-psychology were often quite explicitly formulated. One may suspect that all that makes them count as folk-theories, as against ordinary scientific theories, is that they turned out to be quite drastically misguided!

Does ordinary talk of the mental have even implicit theoretical content? Churchland argues, I think convincingly, that it does. He invites us to compare “numerical” and “propositional” attitudes. A propositional attitude ascription may have the form “X believes that p” or “X hopes that q,” etc. A numerical attitude ascription may have the form “X has a velocity of n” or “X has a temperature of f,” etc. Now just as the embedded numerical attitudes figure in lawful statements quantifying over numbers and expressing contingent relations holding in the domain, so too propositional attitude ascriptions must gain their explanatory force by figuring in (implicit) folk-psychological laws quantifying over propositions and expressing contingent relations believed to hold in the psychological domain. A typical numerical law might be;

\[(x) (f) (m) \left[ (x \text{ has a mass of } m) \text{ and } (x \text{ suffers a net force of } f) \right] \rightarrow (x \text{ accelerates at } f/m)\]

while a typical psychological “law” might be:

\[(x) (p) \left[ (x \text{ fears that } p) \rightarrow (x \text{ desires } -p) \right]\]

(Churchland 1981, p. 71)

Thus suppose we return to our initial example of the postman. Given in detail, our common-sense explanation of the postman’s behaviour might run like this. The postman saw the Alsatian looking angry, and believing that angry dogs cause injuries and desiring to avoid being injured, quickly shut the gate and went on his way. This account is *explanatory* of the postman’s behaviour only if we tacitly accept a general psychological law viz.

\[(x) (p) (q) \left[ (x \text{ sees that } p) \text{ and } (x \text{ believes that } (p \rightarrow q) \text{ and } (x \text{ desires } -q)) \rightarrow (x \text{ will try, other things being equal, to avoid } p) \right]\]

This could be tightened up, but the moral is clear enough. The explanatory force of our ordinary account of why the postman shut the gate depends on our exhibiting his behaviour as falling under a general psychological law. It
is the framework of general laws, said to be implicit in our ordinary explanations of each other’s behaviour, which is meant to constitute the theoretical content of folk-psychology (see e.g., Churchland, 1981, p. 68-69).

I think it is worth stressing the implicit nature of this theoretical content because this may be a clue—though by no means a decisive one—to the status of the theory itself. For theoretical commitments which we tend not consciously to articulate are of two broad kinds. There are those commitments which, though perhaps once explicit, have now become engrained in our talk and culture (e.g., some moral and sexual taboos and practices). And—more interestingly—there are those commitments which we cannot help but make, courtesy of our biological nature and physical environment (e.g., the “commitment” to the use of texture gradients as indicators of surface orientation). The foes of belief-desire psychology would characterise it as implicit in the former, historical and culture-bound sense. But I believe it may be implicit in the latter innately specified sense. Some reasons for choosing this alternative are developed in sections IV and V.

III. DOUBTING THE VOICE OF THE PEOPLE

If folk-psychology does thus involve (implicit) theory, is the theory sound? Stich and Churchland are sceptical. Folk-psychology, they suspect, may prove to be a pandora’s box packed with misguided assumptions and conceptual confusion. Underlying their scepticism we can discern three putative adequacy conditions on a satisfactory psychological theory of the human mind. They are

(1) That the underlying theory should be more than a species-specific guide dealing with behaviourally central cases in normal agents;
(2) That any good theory should exhibit frequent change and expansion;
(3) That there should be some rich isomorphism between neurophysiological details and true psychological descriptions.

All these conditions strike me as being mistaken. The first two suffer from a deep-rooted failure to appreciate the nature and biological function of naive psychological understanding, while the third depends on a failure to recognise the relational (ecologically embedded) content of all good psychological description. Despite the undoubted importance of laying the third ghost to rest, I shall concentrate only on conditions (1) and (2) in this paper, since it is only these conditions which will be directly affected by the alternative account to be developed in Section IV. (But for a good discussion of (3) see W. Bechtel (1985) pp. 473–497.)

Condition (1) surfaces in Churchland’s complaint that the “substantial . . . explanatory and predictive success” of folk-psychology must be set against its failure to cope with
the nature and dynamics of mental illness, the faculty of creative imagination,
...the ground of intelligence differences between individuals...the nature
and psychological functions of sleep...the miracle of memory (and)...the
nature of the learning process itself. (Churchland, 1981, p. 73)

Stephen Stich, in a similar vain, is worried by the failure of folk-psychology to deal with (successfully explain the behaviour of) "exotic folk" and animals. Thus, he claims that we are unable adequately to characterise the content of alien or outlandish beliefs. Citing the example of someone who seems to believe he is a heap of dung, Stich notes that we are tempted to say that if it seems as if someone believes that, then we really can't be sure what they believe. That is, it seems unlikely that any such alien belief can be adequately captured by any ordinary sentence of our language. Another example, again developed by Stich (1983, p. 104) concerns the description of a dog as believing there is a squirrel up an oak tree. In one way, the ascription of such a belief seems fair enough—the dog saw it go up the oak tree and now sits at the bottom waiting for it to come back down. But in another sense it seems quite unwarranted to credit the dog with the belief that what it sees is a squirrel. (Does it also believe it sees an animal? Something that stores nuts?) Stich's point, then, is that in cases of exotic or animal beliefs, folk-psychology shows signs of breaking down. It seems to urge both that someone does and that they cannot believe they are a heap of dung. It seems to urge that the dog does and does not believe there is a squirrel up a tree. Folk-psychology, it seems, it just not up to taking on the really hard cases (Stich, 1983, p. 101). And so much the worse, if Stich and Churchland are to be believed, for folk-psychology as a theory of mental life.

Moving on to Condition (2), we face the accusation that folk-psychology has a history of "retreat, infertility and decadence" whereas a good theory would exhibit progress, refinement, and expansion. The thought here (see e.g., Churchland 1981, p. 74) is that, considered as a standard speculative scientific theory, folk-psychology would seem to be degenerating in the strict Lakatosian (Lakatos 1974, p. 91-196) sense. A scientific theory, or theory-sequence, is said to be degenerating if it fails over a long period to extend its early successes and to predict and explain novel phenomena. Given the implicit nature of the theoretical substratum of folk-psychology it is somewhat difficult to see exactly what the point about degeneration can amount to. Churchland makes his complaint by saying that:

The F.P. (folk-psychology) of the Greeks is essentially the F.P. we use today and we are negligibly better at explaining human behaviour in its terms than was Sophocles. This is a very long period of stagnation and infertility for any theory to display... (F.P's) failure to develop its resources and extend its range of successes is therefore darkly curious and one must query the integrity of its basic categories. (Churchland, 1981, p. 74) (Italics are my emphasis.)
Presumably, then, the thought is that our daily explanations of each other's behaviour (cases like the cowardly postman) ought ideally to be increasing in variety (new terms and phrases) and hence in detailed predictive power and success. This would be evidence of a progressive underlying theory. One immediate comment is that something like this actually does take place. New terms and phrases are coined, and they do seem to bring increased understanding. Terms like "mauvaise foi," "Schadenfreude" and perhaps even Freudian notions of ego and id are cases in point. (I owe these examples to Robert Griffiths.) So if there really is stagnation and infertility it must be located at a much deeper level. And indeed, it would be true to say, as remarked above, that the basic framework of ascribing beliefs and desires as explanations of actions is pretty much unchanged across vast stretches of historical time and geographically distant cultures. But this level of unchanging commonality is, as we shall later see, evidence that what we are dealing with is something rather different to a mere stagnant folk-theory. Beneath all these complaints lies a general sentiment, said to be shared by Stich, Churchland, and Daniel Dennett (see Stich, 1983, Chapter 11, note 10) to the effect that folk-psychology is almost bound to prove deeply misguided. The claim is that:

The very fact that it (folk-psychology) is a folk theory should make us suspicious. For in just about every other domain one can think of the ancient shepherds and camel drivers whose speculations were woven into folk theory have a notoriously bad track record. Folk astronomy was false astronomy and not just in detail... However wonderful and imaginative folk theorising and speculation has been it has turned out to be screamingly false in every domain in which we now have a reasonably sophisticated science. (Stich, 1983, p. 229).

There is something very wrong with this picture. Can we really imagine that our ancestors sat around a campfire and just speculated that human behaviour would be usefully explicable using ideas of belief and desire? Surely not. Some such understanding, though not verbally expressed, seems more likely to be a prerequisite of a highly organised society of language-users than a function of their speculations. Here then, we have hold of the thread which may unravel the case against folk-psychology. In the next section, we shall give it a pull.

**IV. KNOWLEDGE FOR THE SOCIALLY MOBILE**

Is common-sense psychology just bad folk speculation about the mental? I think not. To see why, it may prove helpful to draw a parallel with Hayes' (1979) conception of a naive physics. A naive physics is a body of common-sense knowledge of physical laws and concepts which helps us to get around our everyday world of macroscopic objects. Some kind of knowledge (not
necessarily linguistically formulated) of such concepts and relations as fluid, cause, support, above, below, and beside is vital to a mobile, manipulative being. (A leaping monkey, as Boden (1984, p. 162) points out, must have some kind of grasp of distance, flexibility, support, and so on.) The most vital and basic elements of such a naive physics must be either innately specified (Boden, 1984) cites the "visual cliff" experiments on newborn animals as evidence of some innate grasp of depth) or else must flow directly from the operation of (probably quite specialised) learning capacities trained on the available data. But the point for now is that however we get whatever knowledge of naive physics we have, what we get is an implicit understanding of some very basic, but by no means totally inaccurate, physical principles and laws. If human beings are effective folk-physicists on this level, this is presumably not a result of ordinary campfire speculation on our part. Rather, it is because our cognitive capacities are naturally designed to embody, or else quickly to yield, a rough and ready grasp of whatever physical principles are most important for the success of a mobile, tool-using animal.

What thus goes for a naive physics may, if I am right, go for a naive psychology too. Just as the mobile need to know about support, so must the socially mobile know about the mental states (beliefs, desires, motivations) of their peers. For a sound psychological understanding of others must surely make an important contribution to the overall fitness of a social animal. As Nicholas Humphreys (1983) points out there will always be substantial evolutionary pressure on social animals to become more efficient "natural psychologists." For in the case of such animals the other members of the group are often the single most significant factor (for evolutionary flourishing) in the animal's environment. To take just a single example, consider a recent case study of rhesus macaques (monkeys) (Harcourt, 1985). These animals, in order to prosper, seem to be required to make quite sophisticated judgments concerning the motivational states of their peers. Very briefly, support from a high-ranking female tends to be decisive in combat situations. The likelihood of such support is increased by grooming such females.

Thus, if one macaque sees another groom a high-ranking female, he must try to avoid contests with that macaque in the near future. Some knowledge of the likely behaviour of others (in lending and withholding support) is essential to success. It does not seem unduly generous to describe the knowledge mediating observations (of grooming etc.) and predictions of future behaviour as involving some primitive understanding of the motivational states of other members of the group (see Tennant, 1984a, p. 96; 1984b, p. 178). For a contrary view see Maynard Smith (1984, p. 69). Psychological understanding of one's fellows is, if this right, as important for the success of a social animal as recognising food or predators. No one baulks at an evolutionary account of innate competences subserving our capacities to achieve these latter goals. Why not extend this generosity to the psychological realm? To do so we need do no more than admit that:
Selective advantage will accrue as much to genes that fashion brains that understand mental states as to genes that make brains that have them. Tennant (1984b, p. 177).

We may now distinguish two kinds of human theorising. On the one hand, there is the speculative theorising of alchemy, optics, quantum mechanics, Freudian psychology, and so on. But on the other, there is what might be termed the bedrock theorising which enables us to achieve our most basic goals, such as finding our way around the local environment and roughly understanding the mental states of our fellow animals. So-called "folk"-psychology has its roots in this bedrock theorising about the mental and not, as Stich and Churchland seem to think, in the less well tested higher speculative domains of reason. Notice, then, that we do not deny that folk, or better, naive, psychology has real theoretical content. (This move is tried in Double 1985 but is, to my mind, unconvincing.) The real question is how important this theoretical understanding is to us. Where it is very important indeed (e.g., in naive physics and, I suggest, naive psychology) then we may expect some expertise in that domain to have developed by the normal processes of natural selection. The "hard core" of "folk"-psychology, on this model, is learned by courtesy of an innate competence for understanding others. It is therefore the direct product of a cognitive competence which has undergone the rigorous testing of evolutionary success. This is not the unconstrained campfire speculation we were led to expect.

Bedrock theorising, it must be stressed, is not, for all that, an infallible guide to the truth. A bedrock theory is simply one which is learned or arrived at by the employment of exceptionally well-tested (probably dedicated) cognitive competencies. And evolutionary demands of speed and cost-efficiency may favour the development of competencies which yield theories which are usually (but not always) an approximate guide to the most vital facts about the domain. Bedrock theories, in short, are as unlikely to be absolutely correct as they are to be radically misguided.

Armed with the picture of naive psychology as a bedrock theory of the mental, we may now address the doubts about "folk"-psychology raised in the previous section. These concerned
1) Its apparent lack of fundamental change and development over time (stagnation).
2) Its failure to cope with "exotic cases" and animals (parochiality).

These are facts which stand in need of explanation. The opponents of folk-psychology explain them as the inevitable outcome of bad folk-theorising. The theory of folk-psychology can't develop, they say, because it is fundamentally wrong, and its wrongness is evidenced by its very limited domain of partial success (viz., normal human agents). But we can now turn this line of reasoning on its head. Take parochiality first. Any innate compe-
tence for psychological understanding which humans have developed will of course be geared to understanding the psychological regularities involved in the waking behaviour of normal human agents. For it is in the society of such agents that the organism is being, as it were, groomed to function. Parochiality, then, is par for the course and need not diminish our faith in naive psychology within its intended domain of application. Nor should stagnation pose any great problem. Such lack of change and development is precisely what we would expect of a theory which is formed on the basis of an innately specified competence which, to some extent pre-structures the space of our possible naive psychological understandings. Naive physics offers, I suppose, a static kind of understanding of the world. But it is not a bad understanding because of that. Parochiality and “stagnation,” I conclude, are grist to the naive psychologist’s mill.

V. SO WHAT IF IT’S INNATE?

Belief-desire psychology, I have argued, is a bedrock theory (i.e., one based on an innately specified competence) and hence enjoys a status importantly different to that of the folk-theories with which it is sometimes classed. It remains to address some important reservations concerning what properly follows from the observation of such innateness. One argument (suggested by an unnamed referee of this journal) goes like this. Suppose we grant the plausibility of some innate capacity to make judgments about other organisms’ mental states. And suppose (what may also be challenged) we allow that what is thus innate is doubtlessly well-tested and probably approximately correct within its intended domain of application. Still the step from here to a (provisional, partial) acceptance of common-sense belief/desire psychology may be too big for comfort. For what is innate may just be a tendency, some way or other, to ascribe mental states to our peers. And this leaves open the possibility that the standard, common-sense way of doing so (viz., by ascribing beliefs and desires) is indeed a flawed artifact of our historical period, culture, and language. That is, the belief/desire analysis may still be a mere speculative theory in the true Stich-Churchland spirit.

Against such an objection I would urge two points. The first, which has already been mentioned but bears repeating, concerns the ubiquity of the belief/desire framework. This framework is not an artifact of one time or culture. The detail may vary (as it does with, say, colour perception) but the basic form of the system seems more a part of human nature than of human nurture. The dogged persistence of the central notions of “folk”-psychology across times and cultures is itself a fact in need of explanation. And one way of accounting for it is to refer to the belief-desire framework itself, and not just the tendency to ascribe some mental states or others, back to the operation of a shared and unchanging cognitive competence. Innateness, of course,
is not logically implied by ubiquity. An alternative explanation (suggested by Aaron Sloman) of the (claimed) ubiquity of belief-desire explanation would be that humans command a powerful learning mechanism which, trained on a world in which other agents have beliefs and desires, yields an appropriate "folk"-psychology. The explanation of ubiquity, in that case, would lie with a common world and not a common innate disposition to ascribe beliefs and desires. Even so, it would not be implausible to expect such a basic and important regularity to become "wired-in" over time. And anyway, the "common world" explanation itself suggests a certain realism about beliefs and desires which Stich and Churchland seem committed to denying.

The second point concerns the alleged logical space between an innate tendency to ascribe mental states and an innate tendency to invoke a belief-desire framework. It is possible (though I am by no means certain) that no such space exists. That is, it may be that if we are to view others as exhibiting behaviour governed by mental states at all, then we are forced to use something functionally akin to the belief-desire framework to do so. For consider what looks to be involved in any useful innate capacity to form judgments about others' mental states. Any such capacity must, if it is to be susceptible to the kind of direct evolutionary explanation I have suggested, enable those who have it to predict the behaviour of others on the basis of ideas about their mental states. (If it didn't, then it would be no use as a social navigational tool and hence devoid of selective significance.) But an innate tendency to try to model our peers' internal representations of the world would not, in itself, be of any use at all in predicting behaviour. For knowing how someone thinks the world is is no use unless we have some idea of what they want. Knowing that Fred sees food in a corner is little help in predicting Fred's locomotion unless we also know that Fred wants to eat. Behaviour is thus determined by a subtle interplay between a model of how the world is, on the one hand, and a goal-drive system on the other. If we are in the business of using ideas about other beings' mental states to predict their behaviour, knowledge of both these factors is, as far as I can see, essential. But such twin factors (viz., internal models of states of affairs and goal/drive systems) are functionally isomorphic to the common-sense scheme of beliefs and desires combining to cause actions. If we therefore grant that any capacity to form judgments about the causal (action-governing) powers of others' mental states must involve some functional analogue of twin-factor belief/desire understanding, then it is not unreasonable to expect a proclivity towards such analysis to become part of our natural response to a world full of other intentional systems.

This point bears some relation to Davidson's well-known claims concerning the conditions for the interpretation of a natural language (Davidson, 1973). For the ascription of semantic content to alien utterances, he suggests, likewise requires us to play off, in a global and systematic way, hypotheses
about the beliefs and desires of the beings we seek to understand. Content, he argues, cannot be ascribed through knowledge of either factor alone. It may be that this is true not just of linguistic content but also of the content of causally efficacious mental states. Insofar as we seek to understand the causal role of others' mental states, perhaps we must do so by some kind of twin-factor analysis. Something very like a belief/desire psychology being thus compulsory, it should occasion no surprise if nature responded by making it easy by “wiring” it in. It follows from all this that, in my view, were a society to banish all talk of beliefs and desires, its members, insofar as they were able to understand one another’s mental states at all, would doubtless do so by invoking something functionally akin to the belief/desire scheme. Belief/desire psychology, according to the picture here developed, is not a function of any particular natural language or culture. Natural language may reflect the belief/desire framework. But it is natural necessity which causes us to invoke it.

Someone might grant all this and still deny that folk psychological notions offer a legitimate starting point for cognitive science. The fact that we successfully, and perhaps even inevitably, use a belief-desire framework to facilitate our daily interactions is, it may be urged, no reason to grant such a framework any scientific credibility. Usefulness, which is all that evolution cares about, is not truth, and science aims at truth. It is useful, (to borrow an example from Aaron Sloman), to think of plants as wanting water and seeking light. We can even imagine that our ancestors, in order to farm successfully, were compelled to think of plants as having wants and desires. But this ought not to convince us that plants really do have wants and desires. The latter is a scientific proposition and must be up for grabs whatever the evolutionary and historical facts.

This argument, surprisingly, has my full sympathy. Innateness and usefulness are at best clues to truth and cannot be expected to substitute for it. It could be that the whole project of treating one another as intentional systems is a mere evolved convenience substituting for a proper understanding of the non-intentional well-springs of human action. It could be, but I cannot see how. That is, I cannot see how the kinds of actual and counterfactual human behaviour commonly predicted by, for example, ascribing a desire to eat an ice cream and a belief that ice cream vans congregate in Hyde Park, could even in principle be explained in a completely non-intentional framework (see e.g., Pylyshyn, 1986, pp. 1-48 for a full statement of this argument). It is because I believe that we are compelled, even at the scientific level, to treat human beings (and not plants) as intentional systems that I am inclined to favour the only evolutionarily well-tested, successful theory of intentional systems which we in fact have available, that is, folk-psychology.

The claim, then, is not that “folk”-psychology, having a basis in some innate disposition, must therefore be completely or even approximately correct. It is just that some form of intentional systems theory looks forced upon us as part of any scientific account of human behaviour, and we have,
in "folk"-psychology a reasonably successful and highly tested prototype of such a theory. It seems both unnecessary and unwise to ignore that prototype in seeking, as we should, to construct a more refined version for the purposes of cognitive science. The (alleged) innateness of the belief-desire framework thus speaks only in favour of its being a theory worth considering. Its adherents have after all survived nature's attempts at falsification by premature death. Nature may, of course, have cheated—the theory may be a handy rule of thumb with no scientific validity whatsoever. But often, it is most efficient to play it by the book—to see the cliff edge because the cliff edge is there, and cliff edges are the kinds of things we need to know about. What goes for cliff edges may go for beliefs too. Cognitive science, I conclude, is honour-bound to give belief-desire psychology a chance.

VI. SOCIAL CHEATING, SALLY AND ANNE

I would like to end by reviewing some experimental evidence which seems to support some of the conjectures of the previous two sections. Two sets of results are particularly interesting.

The first involves some work carried out by L. Cosmides (1985). Cosmides compared the performance of normal human subjects on various versions of the Wason selection test where these versions differ in the extent to which successful performance involves reasoning in a "social contract" context. The basic (non-social) form of the test involves giving the subject a rule ("If a card has a vowel on one side then it has an even number on the other") and a set of cards with only one face showing (e.g., "E" "4" "k" "7"). The subject must name all and only the cards which she needs to turn over in order to decide whether the rule is followed. The logical form of the rule is, of course, simply "If p then q" and the faces showing amount to "p" "q" "¬p" "¬q." The correct answer is to turn over the "p" card and the "¬q" card, i.e., in our example the "E" and "7." Ordinary subjects do rather badly on such basic forms of the test. It is well-known, however, that there is a so-called "content effect" which sometimes results in improved performance when the task is laid out in more concrete (contentful) terms (see Wason & Shapiro, 1971). Thus, for example, subjects perform better with a set-up involving the same logical structure but a rule like "If a person is drinking beer than he is over 20" and suitable cards citing age on one side and beer-drinking behaviour on the other (Griggs & Cox, 1982). One natural explanation of this is that it is our greater familiarity with the terms of the test which improves performance. Cosmides' results, however, suggest that a rather different explanation is required. For she reports cases where despite familiarity of subject-matter, no content-effect is observed. Examples include tests using the rule "If I eat salad then I drink water" and others using school and transportation formulations (Manktelow & Evans, 1979).
The important observation which Cosmides makes is that the content-effect seems to arise only when the test is cast in terms involving some kind of "social contract." By this she means any case in which the subjects can represent themselves as trying to find cheaters—individuals who are violating the terms of some agreed social rule. Thus content-effect is observed when the rule is for example, "If a letter weighs 2 oz. then it must have 44 cents postage" but not when it involves for example, the salad example given above.

This enhanced capacity to reason about "social contract" situations is, according to Cosmides, best seen as just one example of a general human talent for engaging in cost-benefit analyses of the kind vital to the survival of social animals in a structured society (cf. the example of the rhesus macaques reported in section IV). The best way to explain our special talents in such regards is, Cosmides suggests (p. 151), to hypothesize innate cognitive structures (what she calls "Darwinian algorithms") specifically adapted to facilitate social forms of understanding. Such algorithms would fix on the intentions, desires, and values of others so as to enable us to make the best deals we can (bargain hard if they really want what you have to offer) and to enforce such deals by spotting "cheaters" (as in the examples cited). Cosmides' results strongly suggest that we are indeed "pre-set" by the evolutionary process so as to become adept "natural psychologists."

The second set of results is due to Baron-Cohen, Leslie and Frith (1985). They argue that at least some forms of childhood autism are best understood as a pathology of a certain cognitive mechanism. This mechanism is claimed to be quite specific and independent of ordinary intelligence as measured by IQ. Its role is to facilitate the ability to form second-order representations.

Autism standardly involves a deep inability to come to terms with the social environment. Such children are often described as "treating people and objects alike" (Baron-Cohen et al., p. 37). This pathology, moreover, is not linked to any more general mental retardation. Some autistic children have normal IQs and conversely, some retarded children (Down's syndrome children) are "socially competent relative to their mental age." To account for such a specific pathology, Baron-Cohen et al. propose an impairment of a computationally specifiable capacity to form second order representations. Such an impairment would explain not only the autistic child's incapacity to conceive of others' mental states (to represent them as representing the world to themselves in a certain way), but also the often observed inability of the autistic child to engage in "pretend play," that is, to imagine they are someone else with different beliefs and desires from their own.
To test the autistic child's capacity to form second order representations (their "metarepresentational capacities") the authors developed a simple test in which success would depend on the capacity to represent someone else as holding a different belief from that of the subject. Very briefly, the child is shown two dolls, Sally and Anne. The dolls are placed in a toy room and Sally is shown putting a marble into a basket while Anne "watches". Sally is taken away. Anne then moves the marble out of the basket and into a box. Sally is put back into the room. The child is then asked; "Where will Sally look for her marble?" and a series of control questions designed to rule out failures of memory and failures of accurate knowledge of present location. Subject to these provisos, any child with a normal grasp of the role of others' beliefs in guiding their actions will expect Sally to look for her marble in the basket where she left it. Normal children and Downs syndrome children respond like this. The autistic child, however, fails to allow for the role of Sally's belief and hence suggests (in 80% of cases) that Sally will look in the box where the marble actually is (Baron-Cohen et al., p. 42). These results, it is claimed,

Baron-Cohen et al.'s results tie up with the classic work of Premack and Woodruff (1978). Premack and Woodruff argued, quite convincingly, that our evolutionary ancestor the chimpanzee is similarly able to understand and predict the behaviour of its fellows by developing a "theory of mind." By this they mean that it imputes to its fellows something very like beliefs and desires. The fact that chimpanzees lack our natural language makes it impossible to uphold the strict claim that what the animals impute to one another corresponds exactly to what we mean by beliefs and desires (see e.g., the Churchlands' response in the same edition of the journal). But the prime functional characteristics of the chimpanzees' "metarepresentations" do seem to justify ascription to them of something like a representation of others' knowledge-states (quasi-beliefs) and current goals (quasi-desires). And this, of course, is just what we would expect if, as we suggested earlier, some kind of twin-factor analysis is indeed forced upon any being who would have a "theory of mind" at all. It is at least possible, then, that the cognitive mechanisms which promote our naive psychological expertise (the mechanisms which explain Cosmides' "social contract" content effect and the dysfunction of which explains the pathology of autism) are directly descended from those which explain the chimpanzees' social competence.

There is, finally, a growing body of work in the field of social cognitive development which suggests, for quite independent reasons, that social and
non-social understanding may be subserved by quite distinct cognitive systems. I shall not review this work here, but see for example, Glick (1978), Gelman and Spelke (1981), and Hoffman (1981).

VII. CONCLUSION

In sum, the experimental evidence of Section VI and our own conjectures (Section IV) add up to a powerful alternative to the Stich-Churchland image of (so-called) folk-psychology as an ordinary speculative theory of the mental. This alternative explains, in an innocent way, the “peculiar” features of folk-psychology which have led some theorists to doubt its ultimate integrity. Such features (viz., its parochiality and “stagnation”) are, we argued, perfectly natural phenomena associated with bedrock (see Section IV) human theorising. They would characterise a naive physics as well as a naive psychology. What such features suggest is not the presence of a degenerating speculative theory so much as the operation of some innate cognitive competence. In these cases, stagnant waters run deep. When we perceive others as believers and desirers (i.e., in employing the “hard core” of folk-psychology which we have labelled “naive psychology”) we may be trusting to the output of a specialised cognitive subsystem as well-tested in its domain as our capacities of sight and hearing. And just as sight and hearing are accepted as broadly accurate sources of data to be refined by physics, so naive psychology should be accepted as a broadly accurate source of data to be refined by cognitive science. The shift of emphasis from total mistrust to provisional and partial acceptance is, I think, crucial. A total and consistent observational sceptic could not trust his senses long enough to formulate a good physical theory. To jettison naive psychology in order to construct a science of the mind would be no less foolish.

REFERENCES


