Two Projects for Understanding the Mind: A Response to Morris and Richardson

NICK CHATER[1] and MARTIN PICKERING[2]
[1]*University of Warwick, Department of Psychology, Coventry CV4 7AL, United Kingdom and*
[2]*University of Glasgow, Human Communication Research Centre, Department of Psychology, 56 Hillhead Street, Glasgow G12 9YR, United Kingdom*[⋆]

Abstract. We respond to Morris and Richardson's (1995) claim that Pickering and Chater's (1995) arguments about the lack of a relation between cognitive science and folk psychology are flawed. We note that possible controversies about the appropriate uses for the two terms do not affect our arguments. We then address their claim that computational explanation of knowledge-rich processes has proved possible in the domains of problem solving, scientific discovery, and reasoning. We argue that, in all cases, computational explanation is only possible for aspects of those processes that do not make reference to general knowledge. We conclude that consideration of the issues raised by Morris and Richardson reinforces our original claim that there are two fundamentally distinct projects for understanding the mind, one based on justification, and the other on computational explanation, and that these apply to non-overlapping aspects of mental life.

Key words: Folk psychology, cognitive science, justifications, causes, computation, knowledge-free, knowledge-rich, problem solving, scientific discovery, reasoning

There appear to be two rather different ways of explaining thought and behavior. The first type is our everyday explanation of each other's behavior in terms of propositional attitudes, such as belief and desire. The second type of explanation attempts to understand the mind in computational terms. These projects are often called folk psychology (e.g., P.M. Churchland, 1989; Fodor, 1987; Stich, 1983) and cognitive science (e.g., Fodor & Pylyshyn, 1988; Johnson-Laird, 1983; Pylyshyn, 1984; Stich, 1983), respectively. But these terms mean different things to different authors, so we shall use the neutral labels "CS" and "FP" below.

The difference between CS and FP appears to run deep. FP explains behavior by providing *justifications* for it. A woman's decision to move house is justified by her belief that her old house is in a dangerous area, that she can afford a new house in a better area, that the housing market is favorable at present, and so on. CS, by contrast, explains behavior in terms of *causes*. Specifically, this causal explanation describes the computational mechanisms underlying the relevant aspect of thought or behavior. If the claim that there are two very different modes of explanation of thought and behavior is correct, then the question arises: Are they compatible? Some theorists have argued that they are not, and that either FP (e.g., P.M. Churchland., 1989; P.S. Churchland, 1986) or CS (e.g., Coulter, 1983) should be abandoned. But other theorists (Fodor, 1975, 1987; Fodor & Pylyshyn, 1988; Pylyshyn, 1984) have argued that they are compatible, and have suggested a particular link

---

[⋆] E-mail:martin@psy.gla.ac.uk; nick.chater@warwick.ac.uk

Minds and Machines **7:** 553–569, 1997.

between the two – that CS is formalized FP[2]. The idea is that the arguments used to justify thought and behavior in FP explanation can be formalized and implemented computationally – and that explaining how this occurs is the task of CS.

All of these viewpoints assume that FP and CS have the same subject matter, and differ over whether they can co-operate, or whether they are in competition. We have recently argued (Pickering & Chater, 1995; henceforth P&C) that this assumption is wrong, and that, instead, the two projects deal with entirely non-overlapping sets of phenomena: It is never possible to explain some aspect of mental life in terms of *both* justifications and computational explanation. So the question of compatibility never arises: FP and CS are about *different things*, not different ways of looking at the same thing. In particular, this implies that CS is not formalized FP.

Morris and Richardson (1995; henceforth M&G) note that our claims are "surprising and significant" (p. 340), but claim that our arguments for them are not correct. They suggest that what we have here called FP and CS represent an unduly narrow definition of cognitive science and folk psychology; and that we have therefore mischaracterized the contrast between them, creating a false dichotomy. In this paper, we respond to these claims, and defend and develop our arguments for the fundamental divide between two styles of explanation of cognition. We first consider some of M&R's general arguments, then address the issue of the scope of each approach, and finally consider whether, as M&R claim, the two projects for understanding the mind are not fundamentally separate, but overlap.

1. General Arguments

The foundation of P&C's paper was the claim that computational explanations (as used by CS) are only feasible for "knowledge-free" aspects of cognition – those that do not draw on a person's general knowledge. This claim was combined with the uncontroversial point that the justifications provided by FP apply only to "knowledge-rich" aspects of cognition – those that *do* involve general knowledge[3] – to yield the conclusion that CS and FP concern non-overlapping aspects of thought.

P&C argued for their claim that computational explanation is not feasible for knowledge-rich processes by drawing on lessons from artificial intelligence, where many researchers have attempted to develop computational models of knowledge-rich aspects of thought. P&C claimed that these attempts typically succumb to two problems: the problem of *right information*, that the knowledge required to reconstruct knowledge-rich justifications is unbounded, as a result of what Fodor (1983) calls the isotropy of common-sense knowledge; and the problem of *right reason*, that knowledge-rich inference is typically *non-monotonic*, and that viable principles for this kind of reasoning have not been developed. We developed these points with extensive arguments in our paper, and referenced previous discussions of these issues (e.g., Chater & Oaksford, 1990, 1993; Oaksford & Chater, 1991, 1993).

M&R do not challenge these arguments for the divide between the domains of FP and CS, which lie at the core of our paper. Instead, they seem to have overlooked them entirely. For example, they note that "P&C offer no reason to think that [the focus of AI on simple, well-defined problems] is anything more than a pragmatic limitation, motivated by the tractability of the problems posed" (p. 341). But our entire section "The AI Experiment" offers such reasons. M&R may not find these reasons convincing, but they provide no arguments against them.

They do, however, put forward three considerations which aim to undermine our claim. The first involves an analogy with biology, the second considers the "partiality" of CS and FP explanation, and the third relates to our discussion of right reason.

An analogy with biology. M&R attempt to counter P&C's arguments by suggesting that AI's focus on simple, well-defined problems, such as checkers (Samuel, 1959), rather than knowledge-rich problems, is comparable to the focus on "suitable model organisms" such as *Drosophila* and *E. coli.* in genetics and evolutionary biology. Biologists can then generalize to more complex organisms; and M&R suggest that "There is no reason to think that research in AI is in principle different from cases like these" (p. 341). But the whole burden of our arguments is that there is a qualitative, not merely a quantitative, shift between the problems that AI can and cannot solve: that some depend on general knowledge and some do not. We argued that because of the isotropy (i.e., interconnectedness) of general knowledge, mental processes which depend on *any* general knowledge thereby depend on *all* general knowledge. But AI has found that formalizing all general knowledge and people's non-monotonic reasoning over that knowledge is completely infeasible. This is why AI has succeeded only in domains where formalizing general knowledge can be avoided entirely. For the same reasons, CS succeeds only for knowledge-free cognitive processes.

Partiality. P&C argued that FP justifications for behavior are founded in a vast and unarticulated groundwork of general knowledge of the world which is common between speaker and hearer. The actual information given in a justification is necessarily partial; the problem of spelling out the entire background for any belief or action is the very problem on which artificial intelligence has foundered. In everyday explanation, the justification for a belief or action need merely be traced back sufficiently far that it makes sense given background knowledge common to both speaker and audience. M&R argue that the computational explanation, the basis for CS, is similarly partial: "What is explicitly used in an explanation is only part of the explanation of a computer's behavior." (p. 343). M&R do not elaborate, so it is not entirely clear what they have in mind. But presumably they mean that the behavior of computers may be explained not just in terms of, for example, the program that they are running, but also by environmental impacts upon them (e.g., whether the room is too hot or damp) and the sources of external inputs they receive (e.g., what gets typed into them), and so on. If this is what they intend, they mistake our point. We entirely agree that explaining the behavior of a computer (e.g., what output it prints, or whether it breaks down), like explaining any other aspect of the physical world, is inevitably partial, because the computer is an "open system" which can be influenced by external factors in arbitrary and unpredictable ways. Alternatively, they may mean that computer users sometimes provide partial descriptions of programs (e.g., "The program attempts to provide medical diagnoses by matching the input to a data base of symptoms which are linked to particular illnesses."). These explanations are of course partial and will depend on the audience's knowledge about computer programmming and the domain of interest. But this version of M&R' s argument is not relevant: It is a criterion of adequate computational explanation that it *can* be made precise (e.g., so that it can be implemented in a computer program). Our argument is that this adequacy criterion on computational explanation can only be met for knowledge-free aspects of cognition, because of the intractable problems of formalizing general knowledge that we discussed above.

To help make our argument clear, note that the contrast we draw is not between the kinds of explanation appropriate for two kinds of objects: people vs. computers. Rather it is between two *styles of explanation*: explanation by giving justifications for a belief or action (FP) and explanation in terms of computational mechanisms (CS). FP explanation is (we would argue inevitably) partial, because it relies on background general knowledge. In contrast, CS explanation, which may be implemented as a computer model of thought, cannot be partial in this sense, because the computer program *has* no prior background knowledge which can fill in the missing aspects of the account. Any knowledge the system has must be explicitly coded as part of the computer program. And making explicit such general knowledge is not possible, because, as we argue, of the isotropy of general knowledge – formalizing one aspect of general knowledge leads immediately to the immense and impossible task of trying to formalize it all.

Right reason. Although, as we have noted, M&R do not provide a direct challenge, they are clearly not convinced by our argument regarding right reason that common-sense knowledge-rich inference is non-monotonic, and that formalizing such inference in computational terms has proved to be entirely infeasible. Their discussion is puzzling, because they appear to endorse our viewpoint while maintaining a rhetoric of disagreement. They note that "Humans are remarkably adept at limiting and focusing on a part of a problem, and are often good judges as to what is relevant to the problem's solution. Humans are sensitive to context in a way that artificial systems are not . . . It is a real problem for cognitive science . . . to devise formal models that have these capabilities." (p. 344). M&R are simply pointing out the problem that we have identified: CS cannot deal with knowledge-rich processes unless the "problem for cognitive science" can be solved (a problem which goes under a number of names: the frame problem (McCarthy & Hayes, 1969); the world knowledge problem; the problem of knowledge representation (Brachman & Levesque, 1985); the problem of inductive confirmation (Carnap, 1952)).

In short, mental processes drawing on *some* general knowledge thereby draw on all general knowledge, by isotropy. So mental processes are either knowledge-rich or knowledge-free, depending on whether they draw on general knowledge or not. But the problems of representing and reasoning with general knowledge are completely intractable to current computational methods. Therefore computational explanation only succeeds insofar as it can avoid these problems, by concentrating on knowledge-free aspects of cognition. FP, by contrast, deals directly and exclusively with knowledge-rich processes. Thus CS and FP deal with non-overlapping aspects of mental life.

2. The Scope of Cognitive Science and Folk Psychology

M&R suggest that we have mischaracterized both cognitive science and folk psychology, and that we have thereby *created*, rather than discovered, a division between them. They suggest that our characterizations, while capturing clear cases of both cognitive science and folk psychology, set the boundaries of each too narrowly.[4] But urging a wider, rather than a narrower, usage of "cognitive science" and "folk psychology" has no substantive implications: It does not challenge our claim that there are two approaches to understanding the mind, and that one deals only with knowledge-free aspects of cognition, while the other deals only with knowledge-rich aspects of thought. It does not matter whether the labels "cognitive science" and "folk psychology" are assumed to refer to subject matters which fall wholly on one or other sides of this divide.[5] The important claim concerns the existence and nature of the divide itself.

Of course, the term "cognitive science" is used in different ways, and can, for example, be taken to include all aspects of the study of mind and behavior, including social psychology, anthropology, and the like. This "umbrella" usage of the term does not aim to pick out a particular style of theorizing about the mind, but rather embraces a disparate set of disciplines which offer complementary and potentially interrelated approaches to understanding cognition (see, for example, Gardner, 1985; Posner, 1989; Wason & Johnson-Laird, 1977, for different broad uses of term)[6]. If so, "cognitive science" would certainly include knowledge-rich territory; for example, it would seek to describe the shared beliefs underlying cultural practices, how these underlie metaphor and other aspects of language use (Langacker, 1987; Lakoff, 1987), and our folk conception of each others minds (e.g., Gopnik, 1993; Perner, 1991; Wellman, 1990). M&R urge that this broad usage is appropriate. We agree that it is perfectly reasonable, but that it is simply not the usage that we explicitly adopted in our paper, where cognitive science was identified with the *computational* explanation of mind – what we here call CS.

Similarly, the term "folk psychology" may be used very broadly. For example, it may be used to incorporate not just explanation in terms of propositional attitudes, but also, for example, any commonly held view about the operation of the mind. According to the broad usage that M&R urge, folk ideas about vision or language learning would deal with knowledge-free processes that are the territory of cognitive science.

But M&R's claims about how we should use terminology have no implications for our claim that there are two fundamentally distinct and non-overlapping approaches to the study of mental life. What matters is the fundamental divide; not whether that divide lines up with the pretheoretic intuitions about how "cognitive science" and "folk psychology" should be used.

At many points in their argument, however, M&R appear to be concerned merely with matters of terminology. They claim that our characterization of cognitive science as the project of explaining mental life in computational terms "seriously limits the scope of cognitive science." (p. 344) and that "Worse, even within cognitive science *as they define it,* they seriously distort the nature of the subject matter and the character of explanation in the field" (p. 344; M&R's italics). The nature of this distortion is not described explicitly (and it is presumably common to other authors, such as Fodor, Pylyshyn and Stich, who use the term in the same way as we do). But whatever this distortion might be, it is be irrelevant to our argument, as it is a purely terminological issue.

In the same way, M&R's arguments in their Section 3 that we define folk psychology too narrowly are beside the point. To take just one example, they argue that common-sense explanations of behavior in terms of emotions which are not propositional attitudes should be included as folk psychology. Those aspects of the emotions that are really knowledge-free should, we argue, form a tractable domain for causal explanation. In fact, there is an extensive literature on the causal basis of emotions, focusing on physiological factors underlying arousal (see, for example, Aggleton, 1992; LeDoux, 1991, 1992), rather than information processing. But emotions also seem to depend on a person's interpretation of their physiological arousal (Schachter, 1964), and this interpretation depends on general knowledge. For example, in a classic study, Schachter and Singer (1962) showed that a uniform physiological change (due to an injection of adrenaline) can have drastically different consequences for emotional state, depending on whether subjects were told about the effects of the injection. In a subsequent interaction with an sullen or frivolous person (a confederate of the experimenters), subjects who knew what effects they would experience behaved normally. By contrast, subjects who did not know about these effects experienced emotional reactions of abnormal intensity, becoming either angry or frivolous. Schachter and Singer's (1962) explanation is that these subjects misattributed their heightened physiological state to strong emotion, and experienced intense anger or euphoria in consequence. Emotions partly depend on subjects' *interpretation* of the cause of their own physiological states, and this justificational interpretation draws upon arbitrary amounts of background knowledge. To this extent, emotions are intractable to causal explanation, and therefore to CS. But knowledge-free aspects of emotion, perhaps concerned largely with level of arousal, can be given a causal explanation (although this may be within the province of physiological psychology and neuroscience (see Izard & Zajonc, 1984) rather than CS). The study of emotion, like many other areas of traditional psychological enquiry, includes both knowledge-free and knowledge-rich processes. We would argue that these different aspects of emotional phenomena must be studied in fundamentally different ways.

3. The Limits of the Computational Explanation of Cognition

We have noted that much of M&R' s argument amounts to a terminological quibble about how to use "folk psychology" and "cognitive science." But they also raise a substantive challenge to our claim that there are two styles of explanation of human thought and behavior and that these apply to non-overlapping aspects of mental life: knowledge-free and knowledge-rich processes. They cite three topics from cognitive science where, they claim, computational explanation encroaches onto the territory of knowledge-rich cognitive processes, thus bridging what we claim to be an unbridgeable divide. These examples come from problem solving, scientific discovery, and reasoning. We now argue that each of these topics breaks into a knowledge-free and a knowledge-rich component (just as we argued for the three domains of vision, memory and language processing, in our original article), and that computational explanation applies exclusively to knowledge-free aspects of these topics.

3.1. PROBLEM SOLVING

Solving problems typically requires general knowledge: We use relevant aspects of our past experience to help us solve fresh problems. But some aspects of problem solving may, nonetheless, be knowledge-free. According to our arguments, cognitive science will only be possible in such areas. The current state of problem solving research suggests that there may be knowledge-free processes which can be studied successfully.

Theories of problem solving in CS have focused entirely on knowledge-free processes. To a large extent, this emphasis begins with the way that cognitive science describes what a problem is. A problem is typically viewed as a "space" of possible states, a set of operators which can traverse the space, possibly subject to certain "path constraints," and the goal is to find a legal path from a specified initial state to a specified goal state (or set of states) (e.g., Newell & Simon, 1972). Finding a suitable path is frequently very difficult, because the number of paths typically increases exponentially with the number of moves made so that exhaustive search is not possible, and it may be difficult to tell whether a particular move takes one closer to, or further from, the goal. But, in the present context, what is really significant is that problems are defined to be closed domains, where the set of states and ways of changing state can be prespecified. This is appropriate for almost all problems which have been extensively studied in this line of research: chess (e.g., Chase & Simon, 1973), checkers (Samuel, 1959), the Tower of Hanoi (Gagné & Smith, 1962), cryptarithmetic (Bartlett, 1958), cannibals and missionaries (e.g., Greeno, 1974), the water jug problem (Atwood & Polson, 1976), or theorem proving in geometry (Anderson *et al*., 1981; Lovett & Anderson, 1994). These closed, formal domains do not appear to involve general knowledge. Problems we face in everyday life (e.g., how to get in to the house having lost the key) do depend on arbitrary amounts of general knowledge (e.g., knowledge of apparently irrelevant information, such as neighbors' holiday plans, may favor or rule out ringing their doorbell to ask for help). Such open-ended, knowledge-rich problems are not addressed by computational models of problem solving. We would argue that they are beyond the scope of computational explanation because they draw on general knowledge. But it is, of course, quite possible that some aspects of mental processes involved in solving such problems are knowledge-free. For example, the structure of the search regime may be explicable without reference to general knowledge. Hence there has been much research concerning whether the problem spaces are searched forwards from the initial state, backwards from the goal, or using some strategy which combines both, such as means-ends analysis (e.g., Newell & Simon, 1972). If knowledge-free aspects of processes can be isolated, then it may be possible to understand them in computational terms.

Thus, computational models (e.g., Anderson, 1993; Atwood & Polson, 1976; Simon & Gilmartin, 1973) similarly entirely avoid reference to general knowledge. To the extent that they embody knowledge of any kind, for example in the heuristics that are used to determine which states or operators are worth considering, they embody knowledge of the closed, formal domains in which the problem is defined.

We conclude that the domain of problem solving provides a good illustration of the fact that computational explanation is feasible only for knowledge-free aspects of cognition.

3.2. SCIENTIFIC DISCOVERY

M&R suggest that scientific discovery is a knowledge-rich domain in which computational explanation has proved to be possible. We would argue instead that progress in providing computational models of knowledge-rich aspects of scientific discovery has been almost non-existent. As with problem solving, what progress has been possible has been achieved in simple closed domains, studying knowledge-free processes.

Scientific discovery is notoriously mysterious from any point of view. For this reason, philosophy of science has traditionally focused on the question of how scientific ideas, once discovered, can be confirmed or disconfirmed (Carnap, 1950, 1952; Popper, 1959)[7]. As we noted in our paper, even this problem of (dis)confirmation is profoundly resistant to a formal treatment, and hence to computational explanation. This is because of the isotropy of knowledge: that a person's entire world-view must be collectively assessed against the entire body of evidence available (see Quine, 1953, and Fodor, 1983, for discussion). But the problem of discovery appears more difficult still, because discovery involves finding theories which are as well-confirmed as possible. Therefore understanding how good theories are discovered seems to presuppose understanding how they are (dis)confirmed.

Computational research on scientific discovery also focuses on closed domains. M&R cite Langley *et al.* (1987) whose research addressed problems such as deriving power laws between small numbers of variables, given observational data. General knowledge that may be relevant to such problem solving is typically not considered; the system works only with a small body of special purpose knowledge and heuristics. More generally, research in a range of fields, including machine learning, artificial intelligence and statistics (Pearl, 1988; Shafer & Pearl, 1990; Shavlik & Dietterich, 1990), has tackled problems of rule discovery and predictions that are analogous to those dealt with in science. But all of this work assumes that a closed domain can be identified, which can be studied independently of general background knowledge.[9]

The picture is the same in the domain of expert systems, which M&R also cite as a counterexample to our claim that computational explanations do not tackle general knowledge. M&R refer to DENDRAL, an early expert system (Buchanan, Sutherland & Feigenbaum, 1969), which identifies organic molecules given information about mass spectrographs and magnetic resonances. This system is successful when general knowledge is not required, but unsuccessful when general knowledge is required. The same holds for other systems that tackle tasks as diverse as mineral prospecting (Duda *et al.*, 1976) and medical diagnosis (Shortliffe & Buchanan, 1975). All of these expert systems succeed to the extent to which general background knowledge is not relevant to their performance.[10] Legal expert systems, by contrast, typically have to take account of some general knowledge, because law is concerned with regulation of everyday life. The programmer must decide by fiat which aspects of knowledge are relevant and which are not (e.g., Ashley, 1990), and to reason only terms of this knowledge. Any attempt to build in additional general knowledge typically forces the domain of the expert system to be drastically reduced (e.g., Branting, 1991). Legal expert systems which are successfully in use are employed only to give an initial judgement, rather than a final decision, and users are required to be aware of, and be potentially able to correct, mistakes due to lack of general knowledge (Dayal *et al.*, 1993)[11] In general, expert systems, in all domains, succeed to the extent that they can avoid drawing on general knowledge, or can artificially restrict the knowledge drawn upon, while still obtaining good system performance.

Note however that our arguments do not imply that the process of scientific discovery, and other reasoning processes which draw on general knowledge, must remain totally mysterious, merely that they cannot be understood in *computational* terms. When scientists explain their processes of discovery they provide *justifica-*

tions in terms of previous beliefs, noting analogies with other domains, experiences with related problems, and knowledge that they obtained from colleagues (see, e.g., Ashall, 1994). Scientific discovery can be explained in *folk psychological* terms, just in the way that other thoughts and actions can be explained. Moreover, the folk psychological explanations can be deepened and enriched from experimental and observational studies of scientific discovery. Insights into scientific discovery may be obtained from studying notebooks and other historical material relating to major scientific discoveries (e.g., Tweney, 1985), studies of laboratory scientists at work (e.g., Latour & Woolgar, 1979) and by experimental studies (Tweney *et al.*, 1981). But, despite M&R, the processes that these explanations describe do not overlap with those that are studied in *computational* research on scientific discovery.

3.3. REASONING

Most human reasoning is concerned with the everyday world, and hence, inevitably draws on general knowledge. But, as with problem solving and scientific discovery, there may be knowledge-free aspects of human reasoning that can be modeled computationally.

Much cognitive psychological research on reasoning has avoided reference to general knowledge, just as research on problem solving and scientific discovery has. This is built in from the beginning in the study of *deductive* reasoning (Evans *et al.*, 1991; Johnson-Laird & Byrne, 1991), because deduction is monotonic, and hence depends only on the stated premises, irrespective of general knowledge. Moreover, many studies of non-deductive reasoning also involve artificial materials to minimize interference from general knowledge (e.g., Klayman & Ha, 1987).

Theories in the psychology of deductive reasoning also typically steer away from general knowledge, and concentrate on the structure of the reasoning system. In deductive reasoning, mental logic theories (e.g., Rips, 1994) assume that reasoning follows natural deduction rules; mental models theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991) assumes that people construct concrete models of the situation described in the premises, read off conclusions from these models, and search for further models which provide counterexamples to those conclusions. Theories involving heuristics and biases (e.g., Evans, 1989) propose that these theories must be supplemented by cognitive biases towards, for example, negative conclusions. These are proposals about the structure of the reasoning system; they make no reference to the knowledge it uses.

The same holds for the tradition in the psychology of reasoning, to which M&R refer, which deals with non-deductive, *probabilistic* reasoning. Proposals in this area range from on formal probabilistic models (Phillips & Edwards, 1966), approximations to these models (Cheng & Novick, 1992; Gigerenzer &: Goldstein, in press) or heuristics and biases (Kahneman *et al.*, 1982). But, as with theories of deductive reasoning, these accounts focus on the structure of the reasoning system and makes no reference to general knowledge (which may be involved, for example, in determining the subjective probabilities from which reasoning begins).

Theorists interested in deductive and probabilistic reasoning do not, of course, deny that general knowledge is *relevant* to human reasoning. General knowledge can, for example, be crucially important in determining factors regarding how a theory of reasoning is applied; but the theory of reasoning itself simply takes those factors as given. For example, consider Oaksford and Chater's (1994, 1995a) model of Wason's (1966, 1968) selection task, one of the reasoning tasks that M&R cite. In this task, subjects are given a rule of the form, *if p then q*, and four cards, each of which have p or not-p on one side, and q or not-q on the other side. The subject can only see the uppermost face of each card. These show p, not-p, q and not-q. The subject is asked which cards should be turned over in order to check whether or not the rule holds.

Two very different patterns of results are observed in different version of this task. In some variants, subjects favor turning the p and q cards; in others, they favor turning the p and not-q cards (see Oaksford & Chater, 1994, for a more detailed analysis). In line with Manktelow and Over (e.g., 1987, 1991), Oaksford and Chater assume that difference between these cases concerns whether the rule used is *indicative* or *deontic*. An indicative rule makes a claim about how things *are*; for example, *if an egg is dropped, then it will break*. A deontic rule states how things *should be*; for example, *if a person drinks alcohol, then that person should be over 21*. The difference between the two becomes apparent when we consider the reaction to cases which violate the rule. Such cases cause indicative rules to be rejected: A dropped and unbroken egg is a *counterexample* to the first rule, which be rejected or revised. But a 20 year old drinker is not a counterexample to the second rule: The person's behavior is viewed as in need of revision, and the rule remains in force. Oaksford and Chater (1994) note that only with indicative rules is it appropriate to view the selection as involving the testing of hypotheses, because it makes no sense to test deontic rules. But it does make sense to attempt to *enforce* deontic rules – that is, to identify violations.

Oaksford and Chater (1994) model the indicative selection task (which includes the standard abstract selection task (Wason, 1966) to which M&R refer) as a problem of selecting cards to gain the maximum expected reduction in uncertainty (in the information-theoretic sense of Shannon & Weaver, 1949). Given certain simple and plausible assumptions, they show that this predicts that people should prefer the p and q cards. By contrast, they model the deontic selection task (e.g., Wason & Johnson-Laird, 1972, cited by M&R) as a problem of maximizing expected utility: They assume that there is a gain associated with detecting violators of the rule, and a small fixed cost representing the "effort" of turning any given card. This predicts that people will choose the p and not-q cards.[12]

The predictions of Oaksford and Chater's theory are derived from a purely formal analysis, using probability theory and information theory. But they *allow* for the influence of general knowledge in two ways. First, and most fundamentally, general knowledge may determine whether a subject treats a rule as indicative or deontic (particularly when linguistic clues to deontic status, such as the presence of the "should" in the rule above, are eliminated). Second, their models involve assumptions about the person's estimate of certain specific probabilities: the probability that the antecedent of the rule is true of an arbitrarily chosen object in the task domain; the probability that the consequent is true; and the estimated prior probability that the rule is true, before any cards have been turned. According to this account, rules with different contents may license different predictions, because they will be associated with different probabilities. The way that the subject comes by these probabilities (i.e., why people find particular statements plausible to a greater or lesser degree) is clearly in general knowledge-rich. But this does not make Oaksford and Chater's theory knowledge-rich; for the theory simply takes the results of knowledge-rich processing as *inputs*, and treats them in a purely knowledge-free way. This conclusion is not, we claim, specific to Oaksford and Chater's account of the selection task, but rather applies quite generally to psychological theories of deductive and probabilistic reasoning.

M&R use the selection task to argue that human reasoning cannot be reduced to formal rules. We would suggest that our own arguments concerning the problems of providing computational models of knowledge-rich processes support this conclusion (see Oaksford & Chater, 1995b, for further discussion). But this is quite consistent with the fact that computational models of reasoning systematically focus on knowledge-free aspects of reasoning.

We have suggested that CS tackles each of the domains discussed by M&R precisely to the extent that it avoids the problem of general knowledge. It remains possible, of course, that processes such as problem solving are wholly knowledge-rich, and therefore that they have no knowledge-free aspects that can be studied using artificial tasks and modeled using formal theories (e.g., Fodor, 1983, adopts this position). If this is so, then the attempts to study problem solving, scientific discovery, and reasoning using computational methods cannot succeed. Researchers in these areas must hope that such a pessimistic conclusion is not correct.

4. Conclusion

M&R argued that we mischaracterize cognitive science, folk psychology, and their relationship, thereby creating an artificial divide between them. In this article, we have stressed that our primary concern is to establish there are two different and non-overlapping approaches to explaining thought and behavior, one limited to knowledge-free, and one limited to knowledge-rich aspects of cognition. We have noted that queries over our (relatively standard) usage of "folk psychology" and "cognitive science" have no implications for these arguments. We then considered the substantive concern that M&R raise, that in three key domains, computational explanation has proved possible for knowledge-rich processes. We have argued that computational explanation is only possible in these domains insofar as it is possible to concentrate exclusively on knowledge-free aspects of these cognitive processes. We conclude that M&R's stimulating discussion does not undermine our claim that FP, which explains by providing justifications, and CS, which explains in computational terms apply to non-overlapping sets of cognitive processes. Moreover, if we follow the terminological tradition of Churchland, Fodor, Pylyshyn and Stich, and identify FP with "folk psychology" and CS with "cognitive science," we can still conclude that cognitive science is *not* formalized folk psychology.

Notes

[1] The order of authorship is arbitrary. Thanks to Ulrike Hahn and Mike Oaksford for valuable discussion of some of the issues discussed in this paper.

[2] These authors express this by saying that cognitive science is formalized folk psychology, because they use "cognitive science" to mean CS, and "folk psychology" to mean FP. The title of our original article was intended to represent a denial of this claim, where the terms are understood in this way. But the range of alternative ways in which the terms "cognitive science" and "folk psychology" can be used can potentially lead to misunderstandings, which is why we have shifted to using the neutral terms FP and CS in this paper.

[3] This is uncontroversial because justifications make reference to beliefs, and a person's beliefs constitute their general knowledge. There is no implication that knowledge must be true or justified, in this context.

[4] In fact, our use of terminology is reasonably standard. On "cognitive science", see, e.g., Johnson-Laird, (1983); Pylyshyn, (1984); Stich, (1983); on "folk psychology," see, e.g., Fodor, (1987); Stich, (1983); P.S. Churchland, (1986).

[5] Indeed, we stressed the point that the way in which these labels are used is not relevant to our argument in our original paper, in an attempt (which was clearly not successful) to head off unnecessary controversy over these terms (Pickering & Chater, 1995).

[6] We note that recent work in "situated" cognition has no implications for the present discussion. This approach puts forward new kinds of computational explanation, and falls within CS (at least, to the extent that it is concerned with human, rather than machine, cognition), while deliberately avoiding problems which involve world knowledge (e.g., Agre & Chapman, 1987; Brooks, 1991; Suchman, 1987).

[7] Indeed, Reichenbach (1938) explicitly argues that philosophy of science must focus exclusively on the context of justification, and avoid the context of discovery of scientific hypotheses, because the latter is not governed by rational principles. Hanson (1958) proposed a "logic" of discovery, but quickly toned down this proposal, in the light of a range of difficulties (Hanson, 1961; see Thagard, 1988 for discussion). Recent work in "cognitive" philosophy of science has revived an interest in the psychological processes underlying theory change and hence discovery, although these are discussed in informal, rather than computational terms (e.g., Darden, 1991; Giere, 1988; Nersessian, 1987). An apparent exception is Thagard's (1988, 1992) work on computational philosophy of science, but Thagard is careful not to claim to have provided a computational account of scientific *discovery*.

[8] Scientific discovery is, of course, frequently treated as mysterious, even from the introspective point of view of the scientists concerned. The justification for a scientific idea may only be apparent in retrospect. Nonetheless, the psychological processes of scientific discovery must in some way be guided by the criteria which are involved in theory confirmation, for otherwise the process of scientific discovery would proceed at the pace of random search.

[9] A rather different research tradition in scientific discovery is formal learning theory (Gold, 1967; Glymour, 1991; Putnam, 1965). This theory models scientific theories in terms of Turing machines, and considers the problem of scientific discovery as the problem of choosing a Turing machine, given a set of given observations. This theory has supported some very general, mostly negative, results concerning the difficulty of scientific discovery, without background knowledge. Note that this theory is not intended to provide a specific mechanism for scientific discovery, because it is couched at far too general a level.

[10] Note that "knowledge-free" is used here in our technical sense, meaning not involving general knowledge. Of course, expert systems involve knowledge, but this is highly specific domain knowledge; and expert systems are feasible to the extent that the domain in question can be isolated from general knowledge, at least to an approximation.

[11] One of the few legal expert systems in use, ASSESS, deals with a highly specialized domain: claims to the government for accident compensation under New Zealand law. Indeed, the use of the expert system was only possible because the relevant law was specifically modified to facilitate computerization. ASSESS makes preliminary judgements concerning accident claims. Appeal to a court is available because the system's judgements are sometimes flawed as a result of its extremely limited knowledge-base (Dayal *et al*., 1993).

[12] Oaksford and Chater (1994b, 1995a) also provide a detailed account of a large part of the data obtained on the selection task over the last 30 years. See Almor and Sloman (1996), Evans and Over (1996), and Laming (1996) for critical discussion, and Oaksford and Chater (1996) for a reply.

References

Aggleton, J.P. (1992), *The Amygdala: Neurobiological Aspects of Emotion. Memory, and Mental Function*. New York: Wiley-Liss.

Agre, P. E. & Chapman, D. (1987), 'Pengi: An implementation of a theory of activity', *Proceedings of the American Association for Artificial Intelligence*, Seattle, PA, pp. 268–272.

Almor, A. & Sloman, S. (1996), 'Is deontic reasoning special?' *Psychological Review*, 103, 374–380.

Anderson, J.R. (1993), *Rules of the Mind*, Hillsdale, NJ: Erlbaum.

Anderson, J.R., Greeno, J.G., Kline, P.J., & Neves, D.M. (1981), 'Acquisition of problem solving skill', in J.R. Anderson, ed., *Cognitive Skills and Their Acquisition*. Hillsdale, NJ: Erlbaum.

Ashall, F. (1994). *Remarkable Discoveries*. Cambridge; Cambridge University Press.

Ashley, K. (1990), *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*, Cambridge, MA: MIT Press.

Atwood, M.E., & Polson, P.G. (1976), 'A process model for water jug problems', *Cognitive Psychology* 8, 191–216.

Bartlett, F.C. (1958), *Thinking*, New York: Basic Books.

Brachman, R.J. & Levesque, J., eds., (1985), *Readings in knowledge representation*. San Mateo, CA: Morgan Kaufman.

Branting, K. (1991), *Integrating rules and precedents for classification and explanation*, Unpublished Ph.D. dissertation, University of Texas at Austin.

Brooks, R. (1991), Intelligence without reason. *AI Memo 1293*, MIT.

Buchanan, B.G., Sutherland, G.L., & Feigenbaum, E.A. (1969), 'Heuristic DENDRAL: A program for generating explanatory processes in organic chemistry', in B. Meltzer and D. Michie, eds., *Machine Intelligence 4*. New York, NY: Elsevier.

Carnap, R. (1950), *Logical foundations of probability*, Chicago, IL: University of Chicago Press.

Carnap, R. (1952), *The continuum of inductive methods*, Chicago, II,: University of Chicago Press.

Chase, W.G., & Simon, H.A. (1973), 'Perception in chess', *Cognitive Psychology* 4, 55–81.

Chater, N., & Oaksford, M.R., (1990), 'Autonomy, Implementation and Cognitive Architecture: A Reply to Fodor and Pylyshyn', *Cognition* 34, 93–107.

Chater, N., & Oaksford, M.R. (1993), 'Logicism, mental models and everyday reasoning: Reply to Garnham', *Mind & Language* 8, 72–89.

Cheng, P.W., & Novick, L.R. (1992), 'Covariation in natural causal induction', *Psychological Review* 99, 365–382.

Churchland, P.M. (1989), *A Neurocomputational Perspective*, Cambridge, MA: Bradford Books/MIT Press.

Churchland, P. S. (1986), *Neurophilosophy*. Cambridge, MA: Bradford Books/MIT Press.

Coulter, J. (1983), *Rethinking Cognitive Theory*. London: Macmillan.

Darden, L. (1991), *Theory Change in Science: Strategies from Mendelian Genetics*, Oxford: Oxford University Press.

Dayal, S., Harmer, M., Johnson, P., & Mead, D. (1993), ‘Beyond Knowledge Representation: Commercial Uses for Legal Knowledge Bases’, *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, Ann Arbor, MI: Association for Computing Machinery.

Duda, R. O., Hart, P.E. & Nilsson, N.J. (1976), ‘Subjective Bayesian methods for rule-based inference systems’, *Proceedings National Computer Conference (AFIPS)*, vol. 15; reprinted in Shafer & Pearl (1990): 274–281.

Evans, J.St.B.T. (1989), *Bias in human reasoning: Causes and consequences*, Hillsdale, NJ: Erlbaum.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993), *Human Reasoning*. Brighton: Erlbaum.

Evans, J.St.B.T. & Over, D. E. (1996), Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, 103, 356–363.

Fodor, J.A. (1975), *The language of thought*, New York: Thomas Crowell.

Fodor, J.A. (1983), *The modularity of mind*, Cambridge, MA: MIT Press.

Fodor, J.A. (1987), *Psychosemantics: The problem of meaning in the philosophy of mind*, Cambridge, MA: MIT Press.

Fodor, J.A., & Pylyshyn, Z.W. (1988), ‘Connectionism and cognitive architecture: A critical analysis’, *Cognition* 28, 3–71.

Gagné, R.M., & Smith, E.C. Jr. (1962), ‘A study of the effects of verbalization on problem solving’, *Journal of Experimental Psychology* 63, 12–18.

Gardner, H. (1985), *The mind’s new science*. New York: Basic Books.

Giere, R. N. (1988). *Explaining science: A cognitive approach*, Chicago: University of Chicago Press.

Gigerenzer, G., & Goldstein, D. (in press), ‘Reasoning the fast and frugal way’, *Psychological Review*.

Glymour, C. (1991), ‘The hierarchies of knowledge and the mathematics of discovery’, *Minds and Machines* 1, 75–95.

Gold, E.M. (1967), ‘Language identification in the limit’, *Information and Control* 16, 447–474.

Gopnik, A. (1993), ‘How we know our own minds: The illusion of first-person knowledge of intentionality’, *Behavioral and Brain Sciences*, 16, 1–14.

Greeno, J.G. (1974), ‘Hobbits and orcs: Acquisition of a sequential concept’, *Cognitive Psychology* 6, 270–292.

Hanson, N. R. (1958), *Patterns of Discovery*, Cambridge: Cambridge University Press.

Hanson, N. R. (1961), Is there a logic of discovery? In H. Feigl & G. Maxwell (Eds.), *Current Issues in the Philosophy of Science*, New York: Holt, Rinehart and Winston, pp. 20–35.

Izard, C.E., & Zajonc, R.B., eds., (1984), *Emotions, cognition, and behavior*, Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1983), *Mental Models*, Cambridge: Cambridge University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991), *Deduction*, Hillsdale, NJ: Erlbaum.

Kahneman, D., Slovic, P., & Tversky, A., eds., (1982), *Judgement under uncertainty: Heuristics and biases*, Cambridge: Cambridge University Press.

Klayman, J. & Ha, Y. (1987), ‘Confirmation, disconfirmation and information in hypothesis testing’. *Psychological Review*, 94, 211–228.

Lakoff, G. (1987), *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Laming, D. (1996), ‘On the analysis of irrational data selection: A critique of Oaksford & Chater (1994)’. *Psychological Review*, 103, 364–373.

Langacker, R.W. (1987), *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.

Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J.M. (1987), *Scientific discovery*, Cambridge, MA: MIT Press.

Latour, B. & Woolgar, S. (1979), *Laboratory life*, Beverley Hills, CA: Sage.

LeDoux, J.E. (1991), ‘Emotion and the limbic system concept.’ *Concepts in Neuroscience*, 2, 169–199.

LeDoux, J.E. (1992), ‘Brain mechanisms of emotion and emotional learning.’ *Current Opinion in Neurobiology*, 2, 191-197.

Lovett, M.C. & Anderson, J.R. (1994), Effects of solving related proofs on memory and transfer in geometry problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 366–378.

Manktelow, K.I., & Over, D.E. (1987), 'Reasoning and rationality', *Mind and Language* 2, 199–219.

Manktelow, K.I., & Over, D.E. (1991), 'Social roles and utilities in reasoning with deontic conditionals', *Cognition* 39, 85–105.

McCarthy, J., & Hayes, P (1969), 'Some philosophical problems from the standpoint of artificial intelligence', in B. Meltzer and D. Michie, eds., *Machine Intelligence 4*, Edinburgh: Edinburgh University Press.

Morris, W.E., & Richardson, R.C. (1995), 'How *not* to demarcate cognitive science and folk psychology: A response to Pickering and Chater', *Minds and Machines* 5, 339–355.

Nersessian, N. (Ed.) (1987), *The Process of Science: Contemporary Philosophical Approaches to Understanding Scientific Practice*, Dordrecht: Nijhoff.

Newell, A. & Simon, H.A. (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.

Oaksford, M.R., & Chater, N. (1991), 'Against Logicist Cognitive Science', *Mind and Language* 6, 1–38.

Oaksford, M.R., & Chater, N. (1993), 'Reasoning Theories and Bounded Rationality', in K. Manktelow & D. Over (Eds.), *Rationality*, London: Routeledge, pp. 31–60.

Oaksford, M., & Chater, N. (1994), 'A rational analysis of the selection task as optimal data selection', *Psychological Review* 101, 608–631.

Oaksford, M.R., & Chater, N. (1995a), 'Information gain explains relevance which explains the selection task', *Cognition* 57, 97–108.

Oaksford, M., & Chater, N. (1995b), 'Theories of reasoning and the computational explanation of everyday inference', *Thinking and Reasoning* 1, 121–152.

Oaksford, M.R. & Chater, N. (1996), 'Rational and irrational analysis of the selection task', *Psychological Review*, 103, 381–391.

Pearl, J. (1988), *Probabilistic reasoning in intelligent systems*, San Mateo, CA: Morgan Kaufman.

Perner, J. (1991), *Understanding the representational mind*. Cambridge, MA: MIT Press.

Phillips, L. D. & Edwards, W. (1966), 'Conservatism in a simple probability inference task', *Journal of Experimental Psychology*, 72, 346–354.

Pickering, M.J, & Chater, N. (1995), 'Why cognitive science is not formalized folk psychology', *Minds and Machines* 5, 309–337.

Popper, K.R. (1959), *The logic of scientific discovery*. London: Hutchinson.

Posner, M. (Ed.) (1989), *Foundations of cognitive science*. Cambridge, MA: MIT Press.

Putnam, H. (1965), 'Trial and error predicates', *Journal of Symbolic Logic* 30, 49–57.

Pylyshyn, Z.W. (1984), *Computation and cognition*, Cambridge, MA: MIT Press.

Quine, W.V.O. (1953), 'Two dogmas of empiricism', *From a logical point of view*, Cambridge, MA: Harvard University Press, pp. 20–46.

Reichenbach, H. (1938), *Experience and Prediction*, Chicago: Chicago University Press.

Rips, L.J. (1994), *The Psychology of Proof*. Cambridge, MA: MIT Press.

Samuel, A.G. (1959), 'Some studies in machine learning using the game of checkers', *IBM Journal of Research and Development* 2, 320–335.

Schachter, S. & Singer, J.E. (1962), 'Cognitive, social and physiological determinants of emotional state'. *Psychological Review*, 63, 379–399.

Schachter, S. (1964), 'The interaction of cognitive and physiological determinants of emotional state'. In Berkowitz, L. (Ed.) *Advances in Experimental Social Psychology*, New York: Academic Press.

Shafer, G. & Pearl, J., eds., (1990), *Readings in uncertain reasoning*, San Mateo, CA: Morgan Kaufman.

Shannon, C.E., & Weaver, W. (1949), *The mathematical theory of communication*, Urbana, II.: University of Illinois Press.

Shavlik, J.W., & Dietterich, T.G., eds., (1990), *Readings in machine learning*, San Mateo, CA: Morgan Kaufman.

Shortliffe, E.H. & Buchanan, B.G. (1975), 'A model of inexact reasoning in medicine', *Mathematical Biosciences* 23, 351–379.

Simon, H.A. & Gilmartin, K.J. (1973), 'A simulation for chess positions'. *Cognitive Psychology*, 5, 29–46.

Stich, S. (1983), *From folk psychology to cognitive science*, Cambridge, MA: MIT Press.

Suchman, L. (1987), *Plans and situated actions: The problem of human-machine communication*, Cambridge: Cambridge University Press.

Thagard, P. (1988), *Computational philosophy of science*, Cambridge, MA: MIT Press.

Thagard, P. (1992), *Conceptual revolutions*, Princeton: Princeton University Press.

Tweney, R.D. (1985), Faraday's discovery of induction: A cognitive approach. In D. Gooding & F. James (Eds.) *Faraday rediscovered: Essays on the life and work of Michael Faraday, 1791-1867*, New York: Stockton, pp. 189–209.

Tweney, R.D., Doherty, M.E. & Mynatt, C.R. (Eds.) (1981), *On scientific thinking*, New York: Columbia University Press.

Wason, P.C. (1966), 'Reasoning', in Foss, B. (Ed.), *New horizons in psychology*, Harmondsworth, Middlesex: Penguin.

Wason, P.C. (1968), 'Reasoning about a rule', *Quarterly Journal of Experimental Psychology*, 20, 273–281.

Wason, P.C., & Johnson-Laird, P.N. (1972), *The psychology of reasoning: Structure and content*, Cambridge, MA: Harvard University Press.

Wason, P.C., & Johnson-Laird, P.N. (Eds.) (1977), *Thinking: Readings in Cognitive Science*, Cambridge: Cambridge University Press.

Wellman, H. (1990), *The Child's Theory of Mind*, Cambridge, MA: MIT Press.