Reply to Lam by Peter Jackson, McDonnell Douglas, Research Laboratory, Artificial Intelligence Group, Dept 225, Building 105, Level 2, RM/PTA1, PO Box 516, St Louis, MO 63166, U.S.A.

Lighthill's division of Artificial Intelligence into three categories—Advanced Automation (A), Bridge Building (B) and Central Nervous System (C)—is of dubious utility, other than to Lighthill himself.

As an example, consider rule-based expert systems. Lam and Lighthill would place these firmly in Category A. Thus placed, their indisputable success cannot be used to defend mainstream AI (Category B) against its detractors.

Yet production rules have a very mixed ancestry. Their mathematical origins can be traced back to Emil Post's canonical systems, but they have also seen application and embellishment in computer science (e.g. in the analysis of algorithms) and linguistics (e.g. in Chomsky's work on formal grammar). Newell and Simon then used this formalism as a basis for modelling human problem solving, and developed techniques of protocol analysis that contributed to today's knowledge elicitation practices. Undoubtedly this work inspired the subsequent efforts at Stanford that led to DENDRAL and MYCIN. Meanwhile, Forgy's work on RETE pattern matching made rule interpreters run in a finite time, and his OPS architecture provided useful conflict resolution strategies that made programs follow more anthropomorphic lines of reasoning. *That* is why XCON and XSEL can handle 20–30,000 rules gracefully, not "the accessibility of high-speed computation".

The point is that rule-based systems owe substantial intellectual debts to advances in mathematics (A), computer science (A), linguistics (C), and psychology (C). More than that, however, they owe a debt to the despised Bridge Builders (B) who put these ideas together and made them work. If Newell and Simon aren't mainstream AI heroes, then who is?

One could repeat this analysis for many other significant developments since 1972, such as neural networks and inductive learning programs. Lam seems determined to misunderstand the synergy of mathematics, computer science, engineering, AI and psychology that gave rise to current work on parallel distributed processing, and it suits his purposes to underestimate its importance. He is also careful to omit any reference to the substantial progress in other areas of machine learning that was made in the 1980s and is now well-documented.

Acknowledging any of these achievements would undermine his assertion that AI is always "assisted", and that this assistance constitutes a redefinition of AI. In fact, the goals of AI have changed surprisingly little in the last 20 years, and the distinction between "pure" and "assisted" AI is a red herring, because the line cannot be consistently drawn. How you would distinguish between "pure" and "assisted" *human* intelligence, given that we all have parents, read books, take courses, accept advice, use machines, etc?

Unfortunately, Establishment "experts" are not well placed to understand interdisciplinary developments. Intellectual disciplines are not created, partitioned or destroyed by Research Councils, but formed by patterns of interaction between individual scientists in an international forum of information exchange. Lighthill's attempt to fashion an emerging research area in his own image revealed his own limitations far more than it revealed the limitations of AI.

AI and Anglo-Saxon Attitudes: a response to Martin Lam by Yorick Wilks, Rio Grande Research Corridor, Computing Research Laboratory, New Mexico State University, Box 30001, Las Cruces, NM 88003, USA.

This is a thoroughly English document, in both the good and bad senses that can be attached to the word: good in its intelligence and lucidity, but it is also parochial, referring to little but English work, even though that is now less central to AI than it was. It is much preoccupied with notions like the Great and the Good, their neutrality and, a cynic would say, their amateurism. It is never

quite clear whether Lam sees Lighthill as one of them, or as a Great Scientist in a neighboring field, and having expertise which would, of course, rob his advice of its charm (in fact, interestingly, he was both at once). There is also an underlying theme of the civil service and its position in a goodwill–neutrality–amateurism space. This is not surprising, given both the author's past profession, and the strong (until recently, virtually complete and total) role that government funding and its vicissitudes have played in the funding of AI in Britain. After reading this paper and having memories of decades jogged, I was heartily thankful that corporate capital is now playing a stronger role in British AI, since it was that funding source in the US that prevented the whole field slipping into the government–military maw. The regret in Britain must be that virtually all the new capital has been from US firms, since Britain has proved unable to sustain a profitable, or indeed any, computer industry, and is now the only large EC country in that position. It is difficult not to suspect that government and civil service policies have had a strong hand in bringing things to this pass.

My initial comments, then, show why I am not at the outset an unbiassed reader of Lam's paper: I cannot see senior British civil servants who have been involved in the high-tech area as neutral beneficent figures; I am more prone to see them as corporately responsible—and I mean nothing personal about Mr Lam here, of course, as I know nothing of his history—for much of the mess, where by the last word I mean a situation with no solid consistent support base for AI. I confess that this prejudice will colour my reactions to the paper. Oh, one more jibe before business; if anything were needed to confirm one's prejudices about the civil service, it is the repeated use in the paper of "euphoria" (appearing in similar contexts to "hype") to dismiss AI's promotion, self- or otherwise. Beware those, I remind myself, for whom "euphoria" is a wholly negative notion; never let them near science or any technical development.

What comments and reactions I have are made on the basis of Lam's paper only, and without any access to the original Lighthill report, though I do not think that important at the level of comment that is appropriate here: the issue is not scholarship but the correctness of a taxonomy of the field and its relation to current research. My first reaction, after being reminded of Lighthill's ABC was to feel that my own position on Cognitive Science might have owed something to it, in spirit at least. To put it simply, I have felt for some time that, although Cognitive Science has been great fun, and enabled one to meet a nicer class of person, there is something inherently self-contradictory about it, since different subjects (AI, philosophy, linguistics, psychology etc) exist precisely because they have different methods of establishing the truth of a claim and, more relevantly, what seems to be the same claim can have its truth established differently by two different disciplines within Cognitive Science and could be said, to that degree, to mean something different within each. That is not an original thought, but just a variant on good old operationalism (after Bridgman) yet its consequences for Cognitive Science are serious: it makes it impossible to conceive of a subject that is some kind of set union of component disciplines, for those are simply not additive if claims are to be understood in terms of practices, experiments and proofs, that may differ arbitrarily between the disciplines.

This distrust of "bridge" subjects suddenly seemed to me, on reading Lam, to have been no more than Lighthill's desire to force what he saw as genuine AI into categories A (Advanced automation) or C (Computer-based studies of the central nervous system) where things were theoretically clear and genuine advances were to be expected. But no, and to my relief, these are not the same point: whereas the division of phenomena between the semi-traditional disciplines is fairly clear even if not perfectly so, the division of AI areas by Lighthill (and followed by Lam) is hopeless and chaotic, as was noted by critics at the time of the original report. The real problem with the classification is that A is engineering and must, almost by definition, encompass everything that is AI, whereas C is science, and the difference between them is not the discipline division I described (as between, for example, linguistics and psychology).

Let me enlarge on this for a moment, before turning to actual details of the research and development of the last 17 years, which is where Lam comes in. I take it for granted that AI is defined, if at all, as a task, not as any particular class of methods for achieving the task. Getting a

computer to interpret, say, visual scenes and accurately locate screw-holes, is an AI task no matter what the method employed. That remains true and important, I believe, even though AI researchers, late at night, will slip into phrases like "well, that wasn't a very AI way of doing it". In the end they cannot be serious unless, and this is important, they are among those in AI who either (a) know in advance what method *must* be used (and believers in predicate logic have certainly been accused of this) or (b) they have no interest in computer models of a task unless they are also the methods currently approved by psychology as being the ones humans use. Both these seem to me tolerable deviances, though ones that AIers must continue to argue against. I have done my share of attacking (a) and here is not the place for more of it. As for (b) it is obvious that we should all be impressed by a successful right–left parser of English and, *if successful enough*, it would cause us to rethink psychological orthodoxy based on superficial phenomena like left–right eye movements when reading English. The last point is not wholly serious, but serves only to remind us that psychology cannot be wholly independent of successful machine models of human performance, no matter what their embedded "theory".

If I am right about this matter then all Lam's phrases like "not directly attributable to AI" (speech recognition); "owing little to AI as generally recognized" (machine translation) and "X was achieved but not by AI methods", "Y was achieved by conventional computer science techniques" are simply impermissible and the product of confusion: to do machine translation is to succeed at an AI task, no matter what the method employed. This fallacy is one of the more substantial ones underlying Lam's paper and is the core of his attack on AI's progress since Lighthill. And, of course, the same trouble was already there in Lighthill, not as these awkward phrases, but in the very basis of the ABC taxonomy. That was not, as Lighthill claimed, a division of all activities into three: for all AI activities are in category A. That goes for every task mentioned in the paper from computer vision to chess to machine translation (and indeed to Lighthill's own work on aircraft landing programs). Some of the models/theories (this is not the place to haggle about those terms) embedded in the programs also have served as the basis of explanatory models in psychological experiments (C). But that is not a division of AI topics, it is a quite different use of a component of an AI theory in a different area.

There is no bridge category (B), although not for Lighthill's reasons (accepted by Lam) that there are AI areas in that category that will not succeed. The category B is necessarily empty: all AI is in A, and C is a different discipline, one that may indeed draw on AI models. If any of this is plausible to a reader, then much of Lam is irrelevant, and worse than Lighthill, because, nearly twenty years later, it has less excuse. So, for example, paradigm cases of C (for Lighthill and Lam) are neural networks for learning, and of B are vision and robotics. Yet these are all AI, and everyone in the field knows it. Lam's whole case is destroyed by his acceptance that there are many products of both types on sale: hence they must be automation (of something) and his remarks at the end of section B show that he has noticed this, yet without seeing its disastrous consequences for his case.

It is difficult to see why so able a man as Lighthill got it so wrong, particularly since he was an engineer in a quasi-AI area who should have understood it much better. On the other hand, perhaps he did understand all too well in a sense: psychology was to be respected as an unknown science and hived off but, as an engineer, he thought that only the engineering techniques he knew would automate the bulk of AI and was repelled by the excesses of symbolic AI and the high-priests of logic in particular. If that was what motivated Lighthill then he should have said so: his report would have then appeared far more sympathetic (to readers like this one, at least) than the byzantine taxonomy he invented to define what he didn't like out of existence. If my speculation is right, then Lighthill (like Dreyfus, curiously enough, a very different critic of AI) would now find connectionism very attractive, as an anti-logical, anti-symbolic approach to AI with strong engineering ancestry and pretensions. The only problem with that is that connectionism may also be plain wrong: there just do seem a lot of human faculties that it is hard to imagine being duplicated without some form of reasoning involved. But this is no place for that discussion either.

Whenever I feel gloomy about AI, I remind myself "*Eppur si muove*" (as Galileo said under

wholly different circumstances): and yet it moves. From the Deep Thought chess program, to the MacnTax discs (do not laugh till you've tried to do US tax returns with and without them), to the new Carnegie Mellon speaker-independent speech recognizer, to machine translations getting steadily better, to expert systems and style correctors and the bits of vision and robotics on sale everywhere, we tend to forget how it moves, partly because so much does not come from approved AI laboratories. But that is another matter entirely!

Another disappointment in Lam's paper which makes it not really a worthy successor to Lighthill's, is that he does not know enough of what is going on in AI and so falls back on the tinkering of the gifted Whitehall amateur. Of course that is *ad hominem* and unfair, but he writes "The breakthrough would come if future translation systems can rely to a greater extent on the concept of probing a deep structure, rather than on a combination of morphological recognition, parsing—i.e. identification of syntax—and semantic methods—but this is far away". That sounds technical but in fact isn't: anyone who knows even a little of the field of natural language processing will know that the quotation is subtly muddled in more than one way (e.g. what could that "deep structure" possibly be in opposition to the all things he opposes it to?). One *could* give sense to the opposition, but nothing Lam writes makes one think that he has. This is not mere nitpicking, since some better level of understanding is vital to a serious discussion of this issue. Lam is also hopelessly Alvey/UK centred in what is now a genuinely world-wide field: he clearly does not know of the Carnegie Mellon speech work I mentioned above, and if he thinks there is no working machine vision, well! I certainly would not take any bets against the ATR speech telephony work at Kyoto based on what Lam says about it!

Some of Lam's discussion is valuable: it is fun to see some tally of where, at the edges of his (in my view wholly flawed) classification, Lighthill has been confirmed or otherwise by subsequent developments. He was right, for example, to be against general problem solvers. Those details alone make the paper worth writing and reading, but my fundamental complaints remain: the author's knowledge and analysis are superficial and that deprives the paper of real weight. Consider, in closing the following: "Diagnostic expert systems would be tiresome and expensive to program and to use—and indeed would lack justification—if they did not incorporate rules and hints provided by a human brain, which enable the system to economize on the volume of search and/or processing". The same muddles are all in there for such systems either work or they do not, they need no other justification. Of course any hints such a system has come from a human brain, where else, but that has nothing to do with whether they *model* a brain (which is what Lam intends but is unknowable in any case). Again, if he means that automation systems of this class need heuristics that express certain human shortcuts (as opposed to "complete" algorithmic solutions) he is just wrong: some do and some do not, and in the end only the market decides what is worth programming, making, selling or using.

Lessons from the Lighthill Flap by John McCarthy, Stanford University.

Martin Lam gives us a British civil servant's view of the Lighthill report and subsequent developments. My comments concern some limitations of this view that may be related to the bureaucratic background of the author—or maybe they are just a scientist's prejudices about officials.

Lam accepts Lighthill's eccentric partition of AI research into Advanced Automation, Computer-based Studies of the Central Nervous System and Bridges in between. This classification was not accepted then and has not become accepted since, because it almost entirely omitted the scientific basis of AI.

AI was not developed as a branch of biology, based on either neurophysiological or psychological observation, experiment and theory. It also is not primarily engineering, although an engineering offshoot has recently been developed. Instead it has been developed as a branch of