Alan Bundy
March 3, 1995

Questions to Ask of AI Research

1 Introduction

There was an interesting forum article in the New Scientist of 17 December 1994[1]. The author tried to classify the questions you might ask of a piece of scientific research into seven types. His examples were all drawn from medical, biological or physical science. At first I hoped the same classification might work for AI — it would just be a question of changing the examplars. On closer inspection, this did not work. His classification all assumed a certain mode of science in which experiments on physical entities were done to establish a cause and effect relationship. So one cold ask, for instance, whether the cause were quantitatively sufficient to account for the effect, or whether cause and effect were correlated in time. Unfortunately, these questions do not seem to transfer to non-physical situations, of the type that arise in informatics.

Nevertheless, the idea of classifying scientific questions is appealing. For a start it would be a big help in training researchers not only to ask good questions of others at seminars, but to ask the same questions of their own work. Such self-reflection might also help improve the methodology of the subject.

In this note I start on such a classification of questions. I don't claim that the types identified here are complete, but I do hope that by being specific gaps in the list will become apparent and can then be filled. Feedback is strongly encouraged.

I will ignore all presentational type questions, *e.g.* questions of clarification of what is being said, pointing out apparent errors or problems with the explanation, *etc.* — except when these are of a methodological nature.

2 The Nature of AI Research

To get started we have to assume a uniform model of scientific activity in AI (or more generally in informatics) research. Previously, I have tried to classify AI research into various types. In practice, a piece of research may simultaneously cover several classifications. I assume that AI is the exploration of a space of techniques and that this exploration takes one of the following forms:

1. Describes a new technique.
2. Extends or improves an existing technique.
3. Establishes properties of a technique or relations between two or more techniques either empirically or theoretically.
4. Describes a new application of a technique.
5. Tests the psychological validity of a technique.
6. Combines several techniques into a system.
7. Identifies and motivates a new task.
8. Establishes properties of a task.
9. Serves a tutorial role, eg a survey.

The methodology usually involves building a computer program embodying the techniques and testing it on some examples. It may just involve a thought experiment about such a program. Or it may involve proving theoretical results about the technique, *e.g.* using mathematics or observing the technique in animals (including humans).

We thus need a different set of questions for each kind of activity. We take these in turn in the sections below.

*Notes in this series are for ϵ baked ideas, for $1 \geq \epsilon \geq 0$. Only exceptionally should they be cited or distributed outwith the Mathematical Reasoning Group.

[1] "Seven pillars of wisdom" Samuel Baron.

3 Describing the Technique(s)

Suppose someone is describing a technique or an extension or improvement of a technique for some set of tasks.

1. What is the technique good for, *i.e.* what tasks will it do? Does it have a wide range of application? Are there applications of practical interest?
2. What are the criteria for success of the technique, *i.e.* how can we know when it is "doing the right thing"? Are these criteria essentially theoretical, empirical or psychological?
3. Is there a characterisation of the tasks (a) to which it is applicable and (b) on which it is guaranteed to succeed (assuming a notion of success exists)?
4. Can the technique be formally described? If so, can properties of soundness, completeness, complexity, *etc.* be shown for it?
5. How does it differ from related techniques, *e.g.* for the same task?
6. What is known about its efficiency? Will it terminate in reasonable time on tasks of practical interest?
7. Has it been implemented and tested? If so, with what results?

4 Building and Testing a System

Suppose someone has built a computer program incorporating one or more techniques and then tested it on some examples.

4.1 The System

We can ask questions about the implementation:

1. Is the implementation complete, or are there unimplemented or untested aspects of it?
2. Is the implementation faithful to the theoretical or abstract description of the technique(s)?
3. Could some of the techniques be incorporated into rival systems? Has this been tried and, if so, with what results?
4. Is the program available for testing by independent researchers? Has it been independently tested and with what results?
5. What facilities are provided for interaction? In particular, what variables can be modified when experimenting with the technique(s)? Do these cover the full range of modifications a tester might want to make?

4.2 The Testing

Or we can ask questions about the testing:

1. What hypotheses were the experiments designed to test and were they supported or disproved?
2. How was it determined that the technique was performing correctly?
3. How can the experimental results be explained? In particular, can we attribute effects noticed in the results to the actions of the technique(s)?
4. How extensive was the testing? Was the test set representative, *i.e.* did it include examples of all kinds? How were the test examples chosen — by the author or independently? Have large examples been tested?
5. Did the test set differ from the development set, *i.e.* was some testing done on previously unused examples after all program development had ceased?
6. Were there any negative results, *i.e.* failure of the technique to terminate, to finish in a reasonable time, or to do a task correctly?

7. How efficient was the program during testing and did the actual performance correlate with any theoretical predictions?

8. Has the program been compared with any rival programs for the same task(s)? If so, what were the results?

9. Did the experiments repeat those previously performed by others and, if so, how do the results compare?

10. Have experiments been conducted in which the technique(s) was systematically modified or even replaced? If so, how sensitive was the program's performance to the precise form or presence of the technique(s)?

11. Were the modifications independent of each other or was there interference between them which made it difficult to draw unambiguous conclusions?

12. Are there some tasks which are inherently hard whatever the technique used to solve them? If so, which features make for harder tasks?

5 Theoretical Results

Suppose someone is proving theoretical properties of a technique or relationships between techniques.

1. Are the theoretic results valid, *i.e.* are the arguments free of gaps, flaws or unjustified claims?

2. Did the theoretical studies address questions of practical importance?

3. Did the theoretical studies answer the questions raised in 3.4 about the soundness, completeness, complexity, *etc.* of the technique(s).

4. How did the properties of the technique(s) studied differ from those of related techniques for the same task(s)?

5. What simplifying assumptions were made about the techniques or the tasks in order to make them amenable to theoretical reasoning? Are these assumptions realistic?

6. Could realistic additional assumptions be made which would simplify the theoretical studies?

7. What practical consequences can we draw from the theoretical results?

6 Conclusion

We have a lot more than the seven question types given by Baron. It could be that I was more imaginative or less discriminating than him, but more likely it reflects the different natures of the fields. Description and implementation of techniques is more like engineering than science, so different criteria apply. Theory is more like mathematics. So we would expect a different range of questions for each different methodological approach.

I have ignored certain types of AI research, *e.g.* psychological experiments, motivation of problems and surveys. Each of these would introduce additional kinds of question, but I did not yet work these out.