Commentary on Anderson

Abstract: *52* words
Main Text: *1040* words
References: *123* words
Total Text: *1258* words

Poppering the Newell test

Niels A. Taatgen
Artificial Intelligence
University of Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
Netherlands
+31 50 3636435
niels@ai.rug.nl
http://www.ai.rug.nl/~niels

Abstract

The Newell Test as it is proposed by A&L has the disadvantage of being too positivistic, stressing areas a theory should cover, instead of attempting to exclude false predictions. Nevertheless Newell's list can be used as the basis for a more stringent test with a stress on the falsifiability of the theory.

The idea of the Newell test is obviously inspired by its illustrious predecessor, the Turing Test (Turing, 1950), and can be considered as an elaboration of the topics that have to be addressed by a theory to make it a plausible basis for an intelligent machine. There is a subtle difference between the two tests: although the Turing Test stresses the fact that the computer should be able to make meaningful conversation, the main point is that the judge in the Turing Test is supposed to do everything possible to expose the computer as a fraud. This aspect of the test is very important, because non-critical discussion partners of the computer can easily be fooled by programs like Eliza (Weizenbaum, 1966; also see Lodge, 1984) and its successors. Analogous to the Turing Test, the Newell test has two aspects: a positivistic aspect, the theory should allow models of all areas of cognition, but also an aspect of falsifiability, the theory should restrict and eventually disallow all "false" models (Popper, 1963). The latter aspect, however, has much less prominence in the Newell test than the former. I would like to criticize this, and argue that the aspect of excluding false models is at least as important, and maybe much more important than permitting true models.

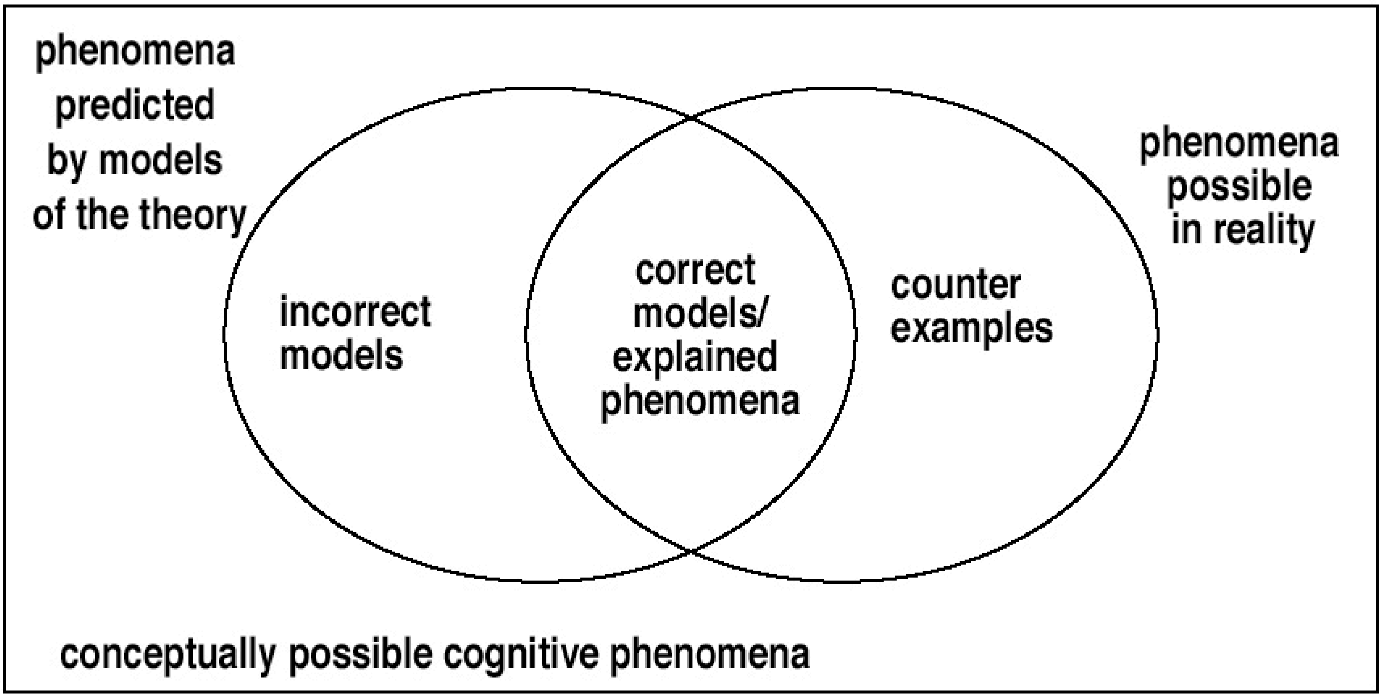

Figure 1. Diagram to illustrate successes and problems of a theory of cognition.

Figure 1 illustrates the issue: consider the set of all possibly conceivable cognitive phenomena, of which only a subset contains phenomena that can actually occur in reality. Then the goal of a theory is to predict which of the conceivable phenomena are actually possible, and the success of a theory depends on the overlap between prediction and reality. The problems of a theory can be found in two categories: counter examples, phenomena that are possible in reality, but are not predicted by the theory, and incorrect models, predictions of the theory that are not possible in reality. The issue of incorrect models is especially important, because an unrestricted Turing Machine is potentially capable of predicting any conceivable cognitive phenomenon. One way to make the Newell test more precise would be to stress the falsifiability aspects for each of the items on the test. For some items this is already more or less true in the way they are formulated by A&L, but others can be strengthened, for example:

Flexible behavior. Humans are capable of performing some complex tasks after limited instructions, but other tasks first require a period of training. The theory should be able to make this distinction as well, and predict whether humans can perform the task right away or not.

Real-time performance. The theory should be able to predict human real-time performance, but should not be able to predict anything else. Many theories have parameters that allow scaling the time predictions. The more these parameters are present, the weaker the theory is. Also the knowledge (or network layout) that produces the behavior can be manipulated to adjust time predictions. Restricting the options for manipulation strengthens the theory.

Knowledge integration. One property of what A&L call intellectual combination is that there are huge individual differences. This gives rise to the question how the theory should cope with individual differences: are there certain parameters that can be set that correspond to certain individual differences (e.g., Lovett, Reder & Lebiere, 1997; Taatgen, 2002), or is it mainly a difference in knowledge people have? Probably both aspects play a role, but is of chief importance that the theory should both predict the breadth and depth of human behavior (and not more).

Use Natural Language. The theory should be able to use natural language, but should also be able to assert what things cannot be found in a natural language. For example, the ACT-R model of learning the past tense shows that ACT-R would not allow an inflectional system in which high-frequency words are regular and low-frequency words are irregular.

Learning. For any item of knowledge needed to perform some behavior, the theory should be able to specify how that item has been learned, either as part of learning within the task, or by showing why it can be considered as knowledge that everyone has. By demanding this constraint on models within a theory, models that have unlearnable knowledge can be rejected. Also the learning system should not be able to learn knowledge that people cannot learn.

Development. For any item of knowledge that is not specific to a certain task, the theory should be able to specify how that item of knowledge has been learned, or to supply evidence that that item of knowledge is innate. This constraint is a more general version of the learning constraint. It applies to general strategies like problem-solving by analogy, perceptual strategies, memorization strategies, etc.

Another aspect that is of importance for a good theory of cognition is parsimony. This is not an item on Newell's list, because it is not directly tied to the issue of cognition, but it was an important aspect of Newell's research agenda. This criterion means that we need the right number of memory systems, representations, processing and learning mechanisms in the theory, but not more. An advantage of parsimony is that is makes a stronger theory. For example, Soar has only one learning mechanism, chunking. This means that all human learning that you want to explain with Soar has to be achieved through chunking, as opposed to ACT-R, which has several learning mechanisms. Of course, Soar's single mechanism may eventually be found lacking if it cannot account for all human learning.

To conclude, research in cognitive modeling has always had a positivistic flavor, mainly because it is already very hard to come up with working models of human intelligence in the first place. But as cognitive theories gain in power, we also have to face the other side of the coin: to make sure that our theories rule out wrong models. This is not only an issue for philosophers of science, but a major issue if we want to apply our theories in human-computer interaction and education. There, it is of vital importance that we can construct models that can provide reliable predictions of behavior without having to test them first.

References

Lodge, D. (1984). *Small world*. New York: Penguin.

Lovett, M.C., Reder, L., & Lebiere, C. (1997). Modeling individual differences in a digit working memory task. In Shafto, M.G. & Langley, P. (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society* (pp. 460-465). Mahwah, NJ: Erlbaum.

Popper, K.R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge.* London: Routledge.

Taatgen, N.A. (2002). A model of individual differences in skill acquisition in the Kanfer-Ackerman Air Traffic Control Task. *Cognitive Systems Research, 3*(1), 103-112 .

Turing, A. (1950). Computing machinery and intelligence. *Mind, 49*, 433-460.

Weizenbaum, J. (1966). ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the Association for Computing Machinery, 9*, 36-45.