Interaction of Deductive and Inductive Reasoning Strategies in Geometry Novices

Kenneth R. Koedinger
John R. Anderson

Psychology Department
Carnegie Mellon University
Pittsburgh, PA 15213
koedinger@psy.cmu.edu

Abstract*

This paper is part of an effort to extend research on mathematical problem solving beyond the traditional focus on formal procedures (both in the classroom and in problem solving research). We are beginning to investigate students' inductive discovery-oriented strategies and the interaction between these and formal deductive strategies. In contrast to typical classroom problems in math and science which demand the application of a learned formal procedure (e.g., prove X), we gave students more open-ended problems (e.g., is X true?) for which the formal deductive procedure is useful, but other, possibly informal or inductive, strategies are also potentially useful. The normative approach for solving these problems, in fact, requires the use of both a deductive strategy, which is definitive only when X is true, and an inductive search for examples, which is definitive only when X is not universally true. When presented with these problems we found that geometry students have some limited facility to perform the deductive strategy (though, less so in this context than when they are directly asked to write a proof) and use a degenerate version of the inductive strategy. Instead of considering multiple examples and looking for a counter-example, students tend to read off the conclusion from the single example (or model) we provided.

Introduction

This paper presents a preliminary investigation into the nature of novice reasoning in geometry. This area is interesting theoretically as it relates to both research on human reasoning biases and research on novice problem solving, particularly in physics. Both areas have identified a number of situations in which humans behave in apparent contrast to normative expectations. Nickerson, Perkins, and Smith (1985) provide a good review of the empirically established human reasoning biases. The research on naive physics has shown that students' prior conceptions of the physical world are often at odds with the more detailed conceptions of the science of physics. Both to add to these bodies of research and for pedagogical purposes, we have begun to explore novice reasoning behavior in mathematics. While the focus in naive physics research has been on students' conceptions, the focus here is on reasoning strategies, both deductive and inductive. In contrast to typical classroom problems in math and science which demand the application of a learned formal procedure (e.g., prove X), we gave students more open-ended problems (e.g., is X true?) for which the formal deductive procedure is useful, but other, possibly informal or inductive, strategies are also potentially useful. We were interested in seeing to what extent students would bring the deductive strategy to bear, to what extent they would use informal strategies, and, in particular, if we could identify biases in these strategies that lead to systematic errors.

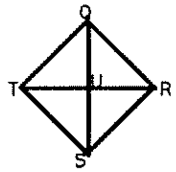
In previous research (Koedinger & Anderson, 1990), we found that skilled geometers initially plan proofs using informal knowledge (perceptually based) and deal with the formal details of writing down the proof as a secondary issue. To better evaluate students' proof abilities in terms of this distinction, we wanted to design a task that tapped only the informal planning abilities and did not require the detailed proof execution skills. The geometry truth judgment task poses questions of the form "If <givens>, must <goal>?". Figure 1 shows an example item.

One approach to answering such questions is to try to find a proof of the <goal> from the <givens>. Subjects don't need to write down this proof, but if they can come up with an accurate one, however informal, they can reliably answer YES.

Besides our initial intention of using this truth judgment task to measure informal planning skills, we have found this task interesting in its own right. From a pedagogical standpoint, it gives students an idea of the role of proof in mathematics, that is, to help determine what is true. From a scientific standpoint, it is analogous to other reasoning tasks and thus, may elicit similar strategies to those observed in everyday reasoning or in other laboratory reasoning tasks. And lastly, it allows us to further explore the complementary roles of informal and formal reasoning strategies.

* Appears in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

11. If $\angle TQU \cong \angle RQU$ and $\angle QTU \cong \angle QRU$, must $\angle STU \cong \angle SRU$?



YES

NO

Figure 1. An example "truth judgment" problem. Subjects circled YES or NO.

A Study of Novice Reasoning in Geometry

Geometry truth judgment problems are much like the logical syllogism problems which have received much attention in the human reasoning literature. A logical syllogism has two premises and a conclusion – these are analogous to our <givens> and <goal>. A typical task is to have subjects judge whether a given syllogism is VALID (must be true) or INVALID (not necessarily true) – analogous to YES and NO, respectively, in the truth judgment task. A leading theory on how humans solve logical syllogisms is Johnson-Laird's (1983) mental model theory. (Polk and Newell, 1989, have refined this theory and implemented it in the Soar architecture.) Essentially, the idea is that subjects convert what is nominally a deductive problem into an inductive problem by thinking in terms of possible models or instances of the premises and testing whether the conclusion holds true for these instances.

A deductive strategy for solving syllogism problems would involve applying logical rules to the premise statements in an effort to derive the conclusion. In contrast, the mental model theory proposes that people pursue a strategy of imagining one or more specific instances which are consistent with the premises. Then they check the problem conclusion to see if it is consistent with the instances they have imagined. In other words, they check whether the conclusion is a proper induction from the instances. If it is, they answer VALID, otherwise they answer INVALID.

What is unique about our geometry problems is that while in typical syllogism experiments subjects imagine their own instance(s), here subjects are given a candidate instance[1]. The diagram that goes along with a typical geometry problem is an instance of the problem premises and, like any model, it has specific features that are not explicitly stated in the premises. Some of these features may actually follow from the premises, for example, $\angle STU \cong \angle SRU$ in the problem Figure 1, while others may be fortuitous, for example, $\angle STU \cong \angle QTU$ in the same diagram.

By giving subjects a model, the truth judgment task provides a different kind of test of the mental models theory. In the typical experiment, the match between the theory and the subjects' actual reasoning process is performed indirectly by comparing the error patterns predicted by the theory with the subjects' error patterns. In contrast, we designed the truth judgment test, in part, to see if we might directly influence the reasoning process (and the resulting answer) by changing the characteristics of the model (diagram) we provide. Figure 2 shows the item design and an example group of 4 matched items varying on whether the correct answer is YES or NO and whether the problem givens and goal "look true" in the diagram or the givens and goal "look false". Consider item b. In this case, the diagram is more specific (has more features) than the given requires and the problem goal $\overline{AB} \perp \overline{CD}$ appears true when, in fact, it does not follow from the given $\overline{AC} \cong \overline{CB}$. In this case, a subject reasoning purely from the provided model would incorrectly conclude the goal was true. In contrast, a subject reasoning purely deductively would not be influenced by this instance, would fail to find a proof and thus, would be led to answer NO (though not with complete certainty since failure to find a proof does not guarantee one does not exist). Of course, the ideal subject would search for a counter-example at this point and upon finding one would be certain that the correct answer is NO.

Item c provides a second kind of misleading or inconsistent diagram. In this case the correct answer is YES yet the goal looks-false in the model (and so does the given necessarily). To the extent that subjects are working from this model, they should be biased to incorrectly respond NO since the goal $\overline{LQ} \perp \overline{RM}$ does not appear true in this diagram. In item types a & d, the model is consistent with the correct answer, so model-based reasoners who rely on this model only will perform better on these item types than on the matched items c & b. Stated another way, subjects who tend to use the provided diagram as a model will answer YES on the looks-true items more often than on the looks-false items.

Our hypothesis was not that subjects would rely purely on the diagram provided. In fact, we were hoping to find evidence of proof planning done in service of solving these problems. While finding evidence that subjects perform above chance on this task would be consistent with the hypothesis that they were bringing proof to bear, it would not rule out other reliable strategies like considering multiple instances of the given statements. However, we can get more discriminating evidence on the use of proof by picking problems for which the underlying proof varies in difficulty. If students are doing proofs, they should be more successful on problems which have "easy" proofs than on problems that require "hard" proofs. We defined difficulty with respect to the conceptual steps in the psychological model of informal proof planning identified by our prior research (Koedinger & Anderson, 1990).

[1]Subjects were certainly free to construct other models of the premises. However, subjects did this rarely, if at all.

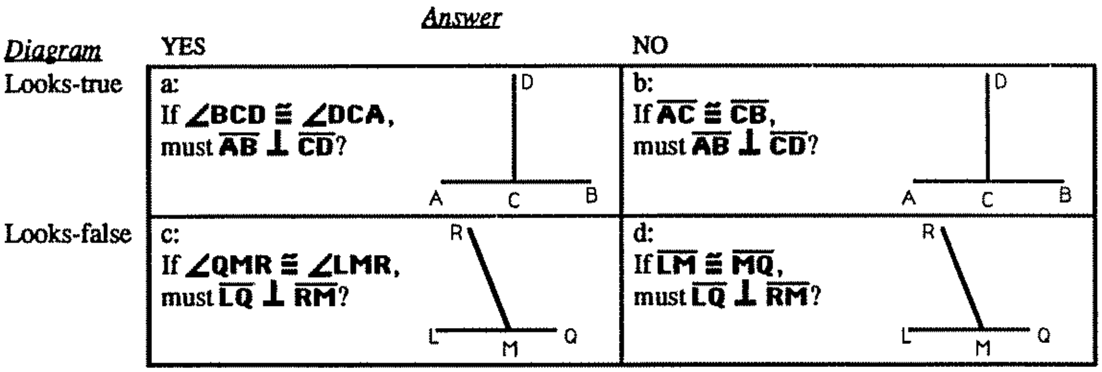


Figure 2. Truth judgment item types. Item types a & c have the same diagram, respectively, as item types b & d, while apart from point labels, a & b have the same problem statement, respectively, as c & d.

The shortest proof that verifies a YES answer on "easy" problems requires one conceptual step, while the shortest proof that verifies a YES answer on "hard" problems requires two or more conceptual steps. This categorization is only directly relevant to the YES items (a & c), but in reporting the data the matching items (b & d) will be included in this categorization.

We designed eight groups (1-8) of four items each (a-d) for a total of 32 items. Item groups 1-4 are "easy" problems, while item groups 5-8 are "hard" problems. The item in Figure 1 is 8a: it requires a "hard" proof, the givens and goal look-true in the diagram, and the correct answer is YES.

Method

Subjects. 30 students participated for pay in the summer *after* completing a high school geometry course.

Procedure. This data was collected as part of the pre and post testing in a study comparing two intelligent tutoring systems (ITS) for geometry proof problem solving (see Koedinger, 1990). The complete set of 32 items was split into two test versions: 16 items for version A and 16 matching items for version B. Half the subjects received version A as a pre-test and B as a post-test; the other half did version B first and A as the post-test. Between the pre and post tests, subjects received 8 hours of proof instruction using one of the ITSs.

The test instructions (1) explained that any points which appear collinear (on the same line) in a diagram can be assumed to be collinear, (2) warned that "*the diagram may be misleading*", and (3) provided two example problems (types c & d) that both contained misleading diagrams and that, in one case, illustrated an informal proof as a good reason to answer YES and, in the other, illustrated a counter-example as a good reason to answer NO.

Results

The first thing to note is that on a separate test of proof skill, we observed a significant improvement in students' ability to write two-column proofs. The mean score went from 36% on a pre-test before the ITS instruction to 68% on the post-test ($p < .001$). Given that the proof problems on the proof test were equally difficult as the proofs underlying the "hard-YES" problems on the truth judgment test, we might expect to see similar improvements on the truth judgment task. However, truth judgment performance showed only marginal improvement, going from 67% on the pre-test to 72% on the post-test ($p > .09$)[1]. While there may have been some transfer, it appears that students were not fully appreciating the relevance of proof to solving truth judgment problems. This is not to say that students were not trying proofs in the service of making truth judgments, in fact, as demonstrated below it is clear they were. Rather, it appears there was some slippage in the transfer from the structured problems in the training where a proof is explicitly required to the open-ended truth judgment problems where students must determine for themselves whether and when proof is relevant. These results are not changed by focussing the data analysis on the items for which improved proof skills might be most relevant (e.g., the hard-YES problems): The improvement trends of about 5% are uniform across all problem types.

Given that there are no qualitative differences between students' performance on the truth judgment pre- and post-tests, the remainder of the analysis uses subjects' scores on the pre- and post-tests combined. The goal is to get a general sense of the reasoning strategies students used to make truth judgments. In particular, to what extent did they bring proof to bear? To what extent are they influenced by the provided diagram? Is there evidence for other strategies? The results are displayed in the left hand graph in Figure 3. This graph shows the mean probability of answering YES for the 8 cells in the 2x2x2 item design. There are three significant main effects on each of three dimensions (correct answer, difficulty, and diagram) and one significant interaction between correct answer and difficulty. They are summarized as follows by indicating which item types lead to a higher probability for answering YES:

---

[1] This improvement is not as quite as small as it seems since getting 50% on these YES or NO problems is not difficult.

1. *Main effect*: YES items > NO items
2. *Main effect*: looks-true items > looks-false items
3. *Main effect*: easy items > hard items
4. *Interaction*: easy-YES > hard-YES, but easy-NO ≈ hard-YES items

All of these effects are significant across both subjects and items ($p < .01$). We will consider them in more detail.

Looking at the answer dimension (represented by the two lines in the graph), students respond YES significantly more often when the correct answer is YES (see the dark line) than when the correct answer is NO (the light line). This indicates that they were not behaving randomly and had some sense for correct answers: they got 66% of the YES-problems correct and 71% of the NO-problems. That proof skill was an important factor in this success is indicated by the fact that performance significantly diminishes going from easy problems (76% correct) to hard problems (61%) and, more importantly, that this difference occurs for the YES-problems but not for the NO-problems. Averaging over diagram appearance, subjects are 83% correct on easy-YES items but 49% correct for hard-YES items while they are 69% correct for easy-NO items and 73% for hard-NO items. In other words, the evidence that students are doing proofs is that students' performance varies in a predictable way on the problems for which proof is required, but not on the problems for which proof is not required.

Though smaller in magnitude, the diagram had a clear and consistent effect on student performance. Independent of the correct answer, students were more likely to say YES on the looks-true items than on the looks-false items. In Figure 3, note how the points drop from the first column (looks-true) to the second (looks-false) and from the third to the fourth. This meant that they made significantly more errors on problem types b & c where the correct answer and the looks of diagram were in conflict than on problem types a & d where the answer and diagram were in accord. Stating this result in a different way, students showed a significant tendency to reason from the provided diagram (or model) making the inductive leap that if the goal looks-true (looks-false) in the model, then YES (NO) it must (not) be true in general.

Modeling Reasoning Strategies and Biases

In this section, we first present a task analysis of what strategies might be brought to bear on this task and the circumstances in which they are effective. With that background, we next discuss what strategies these students appear to have used and how they interact. We propose a mathematical model as a concise summary of our claims about what students are doing. The model is supported both by its close fit to the quantitative data as well as from some preliminary protocol data.

Task Analysis

Effectively solving these truth judgment problems, requires two complementary strategies. A *deductive rule-based* strategy that involves looking for a proof (i.e., a sequence of rules) which derives the problem goal from the givens. If one finds a correct proof, one can conclude YES with certainty. However, if one fails to find a proof, it is not necessarily that case that there isn't one. In this case, one can answer NO heuristically with some better than chance probability of being right. An *inductive example-based* strategy is complementary in that when properly applied one can answer NO with certainty, but can answer YES only heuristically. This strategy, like Johnson-Laird's mental model approach, involves considering examples of the given statements and checking whether the goal is true in these examples. If it is in all cases, one can answer YES but only heuristically since there is no guarantee that the next example won't contradict the problem statement. If the goal is ever found to be false, this example is counter-example to the problem statement and one can answer NO with certainty.

The normative overall strategy is to alternatively pursue one and then the other of these strategies until one of them yields a certain result, that is, until either a proof or a counter-example is found.

A Mathematical Model

By looking at the statistical results from the study, it appeared that students' behavior could be characterized by the interaction of a proof strategy and a degenerate version of the inductive example-based strategy as follows:

1. Subjects initially attempt to find a proof and answer YES if they find one.
2. If they don't find a proof, they "guess" at the answer based on an induction from the provided diagram.

In mathematical terms, this is:

$$\text{prob(YES)} = p + (1-p)g$$

where p = prob of finding a proof
g = prob of guessing YES

This general form is instantiated in different ways depending on the problem characteristics. If the correct answer to the problem is NO, the probability of finding a proof is 0. If the correct answer is YES, the model claims that students find a proof with probability e for an easy problem or probability h for a hard problem. If they don't find a proof, they will guess YES with probability t if the problem goal looks-true in the diagram or f if the goal looks-false. In summary, $p=0$ for NO problems, $p=e$ for easy-YES problems, and $p=h$ for hard-YES problems while $g=t$ for looks-true problems and $g=f$ for looks-false problems.

Note that the model implies that there should be no difference between the easy and hard problems when the correct answer is NO. This is how the model captures the interaction between correct answer and difficulty. The graph on the right in Figure 3 shows the model predictions for the following set of parameter values, $e=.77$, $h=.29$, $t=.37$, $f=.19$. The model provides a close fit to the data. A chi-square test indicates that it does not deviate significantly from the data ($X^2(4, N=30) = 2.19$, $p>.5$)

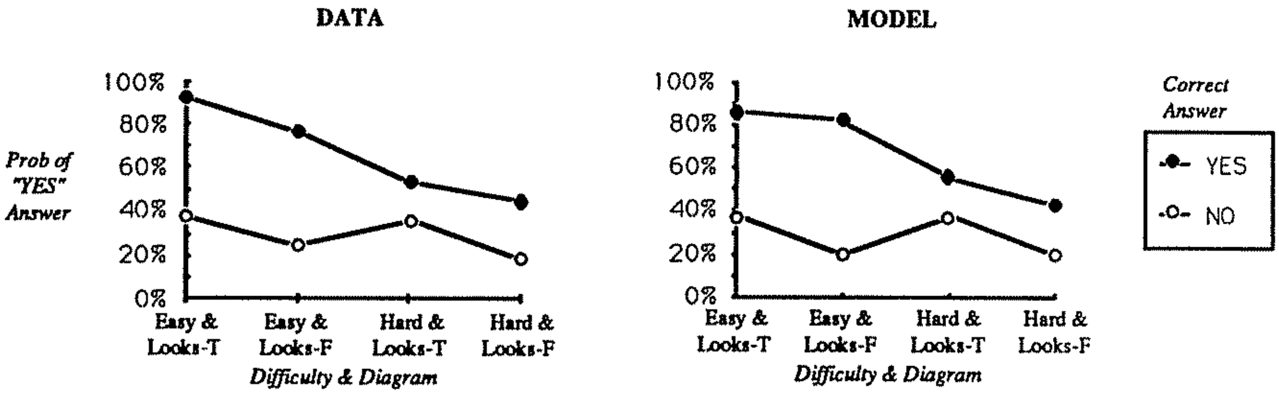


Figure 3. Eight data points and model predictions for the 2x2x2, answer by difficulty by diagram, item design. In the data, all three main effects are significant across both subjects and items and so is the interaction between answer and difficulty. The model does not deviate significantly from the data ($X^2(4, N=30) = 2.19, p>.5$).

This model of student performance suggests three deviations from the normative one. First, the deductive strategy component is quantitatively different in that ideally students should be able to find all the proofs for the YES problems, not just 77% (the value of e) for the easy-YES problems and 29% of the hard-YES problems. Second, the model's inductive example-based strategy is different in that students are not considering multiple examples, but are only considering the provided example and also in that they are not realizing that the looks-false problems are not good examples. Third, there is no alternation in the model between the deductive and inductive strategy.

Other Evidence for the Model

One source of additional evidence for the model comes from the notes students wrote on the test handout provided. In support of the claim in the model that students are not pursuing a search for counter-examples and are not recognizing that the looks-false problems are not good examples, we looked to see how often subjects drew another diagram in addition to the one provided. There were only three cases out of 480 possible opportunities in which a subject redrew a looks-false diagram. While this was a NO problem, none of these diagrams were a counter-example, that is, the goal still looked-true in the diagrams the students drew. Two other instances of redrawing occurred on a different looks-true problem apparently to match up corresponding points of potentially congruent triangles.

We recently gave a version of the test in which we asked students to give reasons for their answers. A large number of the reasons students provide are idiosyncratic, along the lines of "because I think so". But there are also a number of more sensible responses. The reasons students provide for problems they answer YES are often informal proofs, citing intermediate steps and/or theorems from their geometry class. On problems they answer NO, we have yet to find a single instance where the student provided a counter-example. Instead, the most typical, non-idiosyncratic reason for saying NO is that the student fails to find a *proof*.

Conclusion

While these results provide some partial support for Johnson-Laird's mental model theory, there are two key differences from that theory. First, student reasoning was not exclusively model-based (inductive), in fact, more prominent in their reasoning is a deductive strategy to find a proof. Second, while subjects solving syllogism problems appear to consider multiple models, students in this study showed little or no evidence of considering any alternative models beyond the given one. This is particularly surprising given that half of the provided models (the looks-false items) did not accurately represent the problem statement. From a pedagogical standpoint, this study suggests that students need extra instruction both 1) in recognizing the applicability of the proof strategy outside the standard context and 2) in understanding and executing the inductive model-based strategy.

References

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

Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14*, 511-550.

Koedinger, K. R. (1990). Tutoring concepts, percepts and rules in geometry problem solving. Unpublished doctoral dissertation. Carnegie Mellon University.

Nickerson, R. S., Perkins, D. N., and Smith, E. E. (1985). *The Teaching of Thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Polk, T. A., & Newell, A. (1988). Modeling human syllogistic reasoning in Soar. *Program of the Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.