
The Effects of Graduate Training on Reasoning

Formal Discipline and Thinking About Everyday-Life Events

Darrin R. Lehman
Richard O. Lempert
Richard E. Nisbett

University of British Columbia
University of Michigan Law School
University of Michigan

ABSTRACT: *The theory of formal discipline—that is, the view that instruction in abstract rule systems can affect reasoning about everyday-life events—has been rejected by 20th century psychologists on the basis of rather scant evidence. We examined the effects of graduate training in law, medicine, psychology, and chemistry on statistical reasoning, methodological reasoning about confounded variables, and reasoning about problems in the logic of the conditional. Both psychology and medical training produced large effects on statistical and methodological reasoning, and psychology, medical, and law training produced effects on ability to reason about problems in the logic of the conditional. Chemistry training had no effect on any type of reasoning studied. These results seem well understood in terms of the rule systems taught by the various fields and indicate that a version of the formal discipline hypothesis is correct.*

A few years ago an article appeared on the Op-Ed page of *The New York Times* urging that Latin and Greek be taught routinely to high school students in order to improve intelligence (Costa, 1982). The justification given for this recommendation was a study showing that students who had taken Latin and Greek in high school scored 100 points higher on the verbal portion of the Scholastic Aptitude Test (SAT) than students who had not studied these languages.

Although the editors of *The New York Times* apparently thought that this argument was worthy of consideration by its readers, it seems likely that most academically trained psychologists would be dubious on two different grounds. First, because of their methodological training, psychologists would be aware of the likelihood of substantial self-selection effects in any study of the kind described: High school students who take Latin and Greek are likely to be more intelligent than students who do not, and schools that include Latin and Greek in their curriculums are likely to have higher academic standards than schools that do not. Second, most psychologists are aware of the bad reputation of the “learning Latin” approach to teaching reasoning. Thus, they believe reasoning cannot be taught by teaching the syntax of a foreign language, by teaching principles of mathematics, or indeed by any “formal discipline” procedure whereby the rules of some field are taught and then are expected to

be generalized outside the bounds of the problems in that field.

Psychologists are, no doubt, right in their assertion that self-selection undercuts the argument for teaching Latin and Greek. Are they equally justified, though, in assuming that teaching foreign languages or any other formal discipline has no generalized implications for reasoning? The antiformal discipline view not only conflicts with what people have believed for most of recorded history, but its scientific support is far less substantial than most psychologists realize.

The ancient Greeks believed that the study of mathematics improved reasoning. Plato urged the “principal men of the state” to study arithmetic. He believed that “even the dull, if they have had an arithmetical training . . . always become much quicker than they would otherwise have been” (quoted in Jowett, 1937, p. 785). Roman thinkers agreed with the Greeks about the value of arithmetic and also endorsed the study of grammar as a useful discipline for improving reasoning. The medieval scholastics added logic, especially the study of syllogisms, to the list of disciplines that could formally train the mind. The humanists added the study of Latin and Greek, and the curriculum for European education was set for the next several hundred years (Mann, 1979).

The 20th Century Critique of Formal Discipline

Although there were objections to the standard curriculum as early as the Enlightenment on the grounds that the rules of mathematics and Latin bear little actual resemblance to the rules necessary to think about most everyday-life events, it was not until the late 19th century that the view came under concerted attack. The attack came from psychologists, and it was utterly effective. In fact, it was one of the first policy victories of the new field. William James was withering in his critique of “faculty psychology,” that is, of the view that mental abilities consisted of faculties such as memory, reasoning, and will that could be improved by mere exercise in the way that muscles could. Thorndike (1906, 1913) undertook a program of research, still impressive by modern standards, that showed little transfer of training across tasks, for example, from canceling letters to canceling parts of speech or from estimating areas of rectangles of one size and shape to estimating areas of rectangles of another size

and shape. Thorndike declared that "training the mind means the development of thousands of particular independent capacities" (Thorndike, 1906, p. 246). Hence, training in Latin could not be expected to improve people's capacities to perform other very different mental tasks.

Similar conclusions have been reached by contemporary psychologists investigating the transfer of solutions of one problem to solutions of another formally identical problem, for example, between isomorphs of the "Tower of Hanoi" problem (Hayes & Simon, 1977), between homomorphs of the missionaries and cannibals problem (Reed, Ernst, & Banerji, 1974), and between slightly transformed versions of algebra problems (Reed, Dempster, & Ettinger, 1985). Learning how to solve one problem produces no improvement in solving others having an identical formal structure.

The most influential student of reasoning in the middle of the 20th century, Piaget, reinforced already skeptical views of the value of formal training. Piaget believed that there were general rules underlying reasoning—the formal operations and the propositional operations—but that these abstract rules were induced by everyone by virtue of living in the world with its particular regularities (Inhelder & Piaget, 1955/1958). Because learning is mainly by induction, via methods of self-discovery, formal training can do little to extend it or even to speed it up. It is important to note, however, that little research seems to have been conducted examining Piaget's dictum that abstract rules are difficult to teach.

A still more radical view than Piaget's has emerged as a result of studying people's ability to perform certain logical operations. Wason (1966) and other investigators have established that people have great difficulty with abstract problems that follow the form of the material conditional, *if p then q*. This has been done by examining how people respond to selection tasks that embody this logic. For example, subjects are shown four cards displaying an *A*, a *B*, a *4*, and a *7*; are told that all cards have letters on the front and numbers on the back; and are asked to turn over as many as necessary to establish whether it is the case that "if there is a vowel on the front, then there is an even number on the back." Few subjects reach the correct conclusion that it is necessary to turn over the *A* (because if there were not an even number on the back, the rule would be violated) and the *7* (because if there were a vowel on the front, the rule would be violated). More generally, to determine the truth of a conditional statement, the cases that must be checked are *p* (because if *p* is the case, it must be established that *q* is

also the case) and *not-q* (because if it is not the case that *q*, it must be established that it is also not the case that *p*).

Yet subjects have no difficulty solving familiar everyday-life problems formally identical to the Wason selection task. For example, when asked to turn over as many sales receipts as necessary to establish that "if the receipt is for more than \$20, it has a signature on the back," subjects readily understand that large amounts and unsigned reverses have to be checked (D'Andrade, 1982). This fact has led some theorists to argue that people do not use inferential rules at all, but rather they use only those rules that are at a concrete, empirical level (e.g., D'Andrade, 1982; Griggs & Cox, 1982; Manktelow & Evans, 1979). This view would be consistent with the most extreme position derivable from Thorndike's findings: Learning does not transfer from task to task, and subjects do not generalize from a set of tasks to the level of abstract rules.

Pragmatic Inferential Rules

Recently, we and our colleagues have argued that the Thorndike tradition is mistaken in asserting the extreme domain specificity of all rule learning. We have identified several naturally occurring inferential rules that people use to solve everyday-life problems and have found that they are improvable by purely formal training (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986; Fong, Krantz, & Nisbett, 1986; Holland, Holyoak, Nisbett, & Thagard, 1986; Nisbett, Fong, Lehman, & Cheng, 1987). These "pragmatic inferential rules" capture recurring regularities among problem goals and among event relationships that people encounter in everyday life. They are fully abstract in that they are not tied to any content domain (much like Piaget's formal operations), but they are not as independent of relationship types and problem goals as formal logical rules (which are included in Piaget's propositional operations) or the purely syntactic rule systems often studied by modern cognitive psychologists.

Contractual Schemas

One type of pragmatic inferential rule system we have studied we call "contractual schemas." These schemas represent situations in which a *permission* is required to perform some action or in which an *obligation* is incurred by virtue of the occurrence of some event. These schemas are of particular interest because the procedures needed to establish whether a permission or obligation has been violated are the same as the checking procedures required by the conditional to establish whether a proposition of the form "if *p*, then *q*" obtains.

Cheng and Holyoak (1985) showed that the schema for permissions is useful in performing selection tasks having the form of the Wason card problem. In one of Cheng and Holyoak's experiments, subjects were presented with a selection problem based on an abstract description of a permission situation: "If one is to take action 'A,' then one must first satisfy precondition 'P.'" About

We are indebted to Rebecca Collins, Elissa Wurf, and Dara Markowitz for advice and assistance and to Richard Catrambone, Patricia Cheng, and Keith Holyoak for comments on an earlier version of this article. The research was supported by NSF Grants SES85-07342, BNS84-09198, and BNS-8709892, Grant 85-K-0563 from the Office of Naval Research, and Grant 87-1715 from NSERC.

Correspondence concerning this article should be addressed to Darrin R. Lehman, Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, B.C., Canada V6T 1Y7.

60% of the subjects solved this problem as compared with 20% who solved the original Wason selection task. Cheng and Holyoak also found that providing an explicit purpose for a rule that would otherwise seem arbitrary could serve to cue the permission schema and hence facilitate performance. These findings are incompatible with the extreme domain specificity view stemming from Thorndike's position. The findings also indicate that the selection task, with its arbitrary elements and relations, is difficult because people normally reason using schemas of the permission type rather than the rules of formal logic. When reasoning schemas are invoked, people can solve problems that are formally identical to the card selection task because the schematic rule structures are identical.

It also turns out to be the case that training in formal logic has little effect on people's ability to solve either arbitrary or semantically meaningful versions of the selection task, whereas training in the obligation schema has a substantial effect. Cheng, Holyoak, Nisbett, and Oliver (1986) found that neither an intensive training session on the nature of conditional rules nor even an entire course in logic improved subjects' performance. In contrast, Cheng et al. found that even brief instruction in the obligation schema improved subjects' abilities to solve the selection task, especially more semantically meaningful versions of it that could be understood in terms of the obligation notion.

Causal Schemas

Cheng, Nisbett, and Oliver (1987) have argued that another type of pragmatic reasoning schema, namely causal schemas of the kind defined by Kelley (1971, 1973), may also help people to solve problems that are syntactically identical to the selection problem. Kelley argued that people understand the ideas of necessariness and sufficiency in causality and possess different schemas for checking evidence supporting causal hypotheses that are necessary and sufficient, necessary but not sufficient, and so on. Cheng et al. (1987) found support for this view. They found that subjects tacitly assumed a particular type of causality and then applied evidence-checking procedures appropriate to the type. Interestingly, one of the schemas, namely that for sufficient but not necessary causes, has checking procedures identical to that for the conditional selection task. A hypothesis of the type "A among other things always causes effect B" can be disproved by examining A and finding that not-B was the case, or examining not-B and finding that A was the case. Cheng et al. (1986) found that subjects performed better on selection tasks that had a causal interpretation that might encourage subjects to use the same checking procedures as required by the conditional. The checking procedures for the necessary and sufficient type of causality, as it happens, are the same as for the biconditional (*p if and only if q*) in formal logic. The biconditional requires examining all four cases—*p*, *not-p*, *q*, and *not-q*. The other two causal schemas—namely, necessary but

not sufficient, and neither necessary nor sufficient—also have distinct checking procedures associated with them. The social sciences, it should be noted, have developed elaborate methodologies for dealing with the completely probabilistic type of causality characterized by the neither necessary nor sufficient type.

Statistical Rules

Another of the inferential rules we have studied is comparable to Piaget's "probability schema." In our view, this is not a single schema but a family of related schemas or heuristics having to do with the law of large numbers; the rule that sample values resemble population values as a direct function of sample size and an inverse function of population variability; and the related regression or base rate principles, for example, the rule that extreme values for an object or sample are less likely to be extreme when the object is reexamined or a new sample observed. We have found that people often use the law of large numbers and relatively simple applications of the regression principle when solving problems in everyday life (Nisbett, Krantz, Jepson, & Kunda, 1983). We have also found that people's solutions of everyday-life problems using statistical rules are greatly enhanced not only by instruction in college statistics courses but even by relatively brief training sessions (Fong, Krantz, & Nisbett, 1986). These training sessions are effective even when the training is highly abstract and formal and does not make any reference to everyday-life content. In addition, training in one domain of events improves reasoning for other quite different domains fully as much as for the trained domain, suggesting that subjects readily abstract what they learn from a given domain (Fong et al., 1986; Fong & Nisbett, 1988).

Graduate School and Formal Discipline

The work done to date indicates that people reason using inferential rules at a fairly high level of abstraction and that their ability to use such rules can be improved by formal training procedures. Theorists prior to the 20th century thus were probably correct about the basic notion of formal discipline, although they probably misidentified the particular types of rule systems that are amenable to training and that allow substantial generalization to everyday-life problems. The syntactic rule systems of mathematics and formal logic may be destined to play little role in everyday reasoning even when these have been well-taught; pragmatic rule systems, such as the law of large numbers, causal schemas, and contractual schemas, may play a significant role even prior to formal training.

The work also has some clear implications for understanding the effects of education on the way people reason. Because graduate training in particular is highly specialized with respect to the inferential rules emphasized, the possibility exists that different kinds of graduate training produce different effects on reasoning about various everyday-life events. In particular, we would expect that scientific disciplines teach different rules than non-

scientific disciplines and that there might be differences within the sciences themselves having to do with whether the field is at base probabilistic or deterministic. We studied two probabilistic sciences, namely psychology and medicine, a nonprobabilistic or deterministic science, namely chemistry, and a nonscience, namely law.

One would expect that training in the probabilistic sciences would affect statistical reasoning, sensitizing people, for example, to the riskiness of inferences from small samples. One would also expect such training to affect causal reasoning because accurate causal judgments usually require some understanding of the problems posed by confounded variables, the bane of the probabilistic sciences. A methodological education in the probabilistic sciences sensitizes people to problems that arise when studying causes that are neither necessary nor sufficient. It includes rules about when and how to employ control groups, how to avoid sample bias, and how to recognize and avoid the errors that can arise when subjects who are selected or who select themselves for a treatment on the basis of one variable also differ on other variables that are potentially correlated with the dependent variable of interest. In contrast, one might expect that training in a nonprobabilistic science such as chemistry would have less effect on statistical reasoning and on methodological reasoning, which requires sensitivity toward the problems posed by confounded variables. One might also expect training in a nonscience such as the law to have relatively little effect on such kinds of reasoning.

The probabilistic sciences might be expected to improve reasoning on conditional and biconditional problems because these fields distinguish the nature of causality implied by their hypotheses, often in a quite self-conscious fashion. Most hypotheses examined by social scientists are of the neither necessary nor sufficient type. Some hypotheses are of the necessary but not sufficient kind (for example, that only certain kinds of subjects are "at risk" for certain outcomes, although other causes are required for the risk to be manifest). Other hypotheses are of the sufficient but not necessary kind (that is, many different causes may produce a given effect, for example, retardation, although none is required). To the extent that people use these schemas to solve problems for which a logician would use the conditional, we might expect that training in the probabilistic sciences would improve conditional reasoning. Again, one might expect training in a nonprobabilistic science such as chemistry to affect such reasoning less because it focuses primarily on deterministic causality of the necessary and sufficient kind.

We might also expect that training in the law would improve conditional reasoning. The law deals with contractual relations that have the form of the conditional such as permissions and obligations as well as with contractual obligations that have the form of the biconditional, that is, agreements that a certain action may be taken if and only if some event has occurred.

We investigated the effects of different types of graduate education on answers to problems that we thought would be differentially affected by training. As a control,

we examined the effects of different training on verbal reasoning of the kind tested by the Graduate Record Examination (GRE). Verbal reasoning ability increases steadily during the young adult years for those involved in mentally challenging work (Share, 1979). We expected that subjects in all four disciplines would improve slightly, and to about the same extent, in verbal reasoning.

The first study was cross-sectional, examining first- and third-year students enrolled at the University of Michigan in the fall semester of 1983. The second study was a longitudinal one that reexamined the first-year students in the cross-sectional sample at the beginning of their third year.

Study 1: Cross-Sectional Design

Subjects

Subjects were first- and third-year graduate students in law, medicine, psychology, and chemistry at the University of Michigan. Because of substantial attrition in the relatively small chemistry program, the "third-year" chemistry group also included all available fourth-year students.

Many educational studies suffer low response rates or high drop-out rates, raising the possibility that any effects reflect selection biases rather than educational effects of interest. The present investigation, except for the chemistry group, suffers very little from these problems. Almost all the enrolled students in the four programs responded, and the drop-out rates were extremely low in all programs but chemistry. Moreover, as will be seen later, the results for chemistry suggest that the drop-out rate for that group does not pose serious problems for interpretation.

The first year response rates were as follows: 213 of 241 law students (88%), 127 of 133 medical students (96%),¹ 25 of 27 psychology students (93%), and 31 of 32 chemistry students (97%).²

Due to the different numbers of students in the four disciplines and the cost of sampling third-year students, random samples of 60 third-year law students and 55 third-year medical students were contacted. Samples were based on the populations initially enrolled, so that some of the sampled students had dropped out by the time the study began. All third-year psychology students and third- and fourth-year chemistry students were sampled because these programs were relatively small. The third-year response rates were as follows: 50 of 60 ever-enrolled law students (83%, but the figure for currently enrolled students was 89%), 48 of 55 ever-enrolled medical students (87%), 33 of 36 ever-enrolled psychology students (94%), and 26 of 52 ever-enrolled chemistry students (50%, but the figure for currently enrolled students was 93%).

¹ An additional 62 first-year medical students were given a different test for the purposes of another study. The total class *N* was 195.

² Students with English language problems were omitted throughout. This exclusion procedure significantly affected only the chemistry samples. Approximately 18% of chemistry students were excluded as compared to fewer than 3% of those in the other three disciplines.

Procedure

First-year subjects were given a reasoning test during separate mass orientation meetings at the beginning of their first term in graduate school. Third-year subjects were given the test in one of three settings: group administrations in their home department, in their own offices, or in offices at the Institute for Social Research. In all settings, a relaxed, unhurried atmosphere prevailed. Subjects were instructed to take as much time as necessary to solve each problem: "Please treat each fully; but there is no need to puzzle unduly over any problem. While some questions have right and wrong answers, others do not."

Two forms of the reasoning test were developed so that pre- and posttest versions would be available for the longitudinal study. Approximately half of the subjects in each group filled out Form 1 of the reasoning test, and the other half filled out Form 2. The test took between 40 and 60 minutes to complete. The third-year students were paid \$10 for their participation.

Instrument

The reasoning test consisted of four sets of questions, spread evenly throughout the test booklet. One set was intended to measure statistical reasoning; one was intended to measure confounded variable (or methodological) reasoning; one was intended to measure reasoning about problems to which conditional logic could be applied; and one was a standard measure of verbal reasoning of the kind employed in the Graduate Record Examination verbal test.

Statistical reasoning. The statistical reasoning items were intended to measure the subjects' abilities to apply the law of large numbers and the regression or base rate principle to both scientific problems and everyday-life problems. Scientific problems were those that specifically mentioned a study of some kind within the body of the question, whereas everyday-life problems did not. In one of the scientific problems, for example, subjects were told about a teaching experiment using high school students. They were asked what would be expected to happen to the grades of students in a control condition who had relatively high grades and those who had relatively low grades. A recognition of the regression principle was shown by indicating that the high group could be expected to have lower grades in the subsequent term and the low group could be expected to have higher grades. An everyday-life problem to which the law of large numbers could be applied was identical to one used by Nisbett, Krantz, Jepson, and Kunda (1983): A high school student had to choose between two colleges. The student had several friends, who were similar to himself in values and abilities, at each school. All of his friends at school A liked it on both educational and social grounds; all of them at school B had deep reservations on both grounds. The student visited both schools for a day, and his impressions were the reverse. Subjects could indicate their understanding of the law of large numbers by stating that the student should probably choose A because his "one-

day visit can't give him a very good idea of what the school is like, whereas his friends' long-term experience can be informative." An example of an everyday-life problem to which the regression principle (and perhaps the law of large numbers as well) could be applied is presented in its entirety in Appendix A at the end of this article.

Methodological reasoning. The methodological problems were intended to measure ability to apply various confounded-variable principles to problems having both scientific content and everyday-life content. One of the scientific content problems, for example, presented an assertion that the bald eagle population was on the rise based on the fact that a study showed that sightings of bald eagles by the North American Wildlife Federation's annual watch had increased by 35%. An ability to apply control group concepts was indicated by recognizing that increased sightings of other kinds of birds would undermine the claim (because it would suggest more people involved in the watch, different reporting methods, etc.). Another of the scientific problems presented the claim that students who learned Latin or Greek got higher SAT verbal scores and measured subjects' recognition of the self-selection principle. One of the everyday-life content methodological problems, showing recognition of the need for control groups, is presented in Appendix A.

Conditional reasoning. Each form contained three problems that could be solved by application of the conditional and one that could be solved by application of the biconditional. The problems were those used by Cheng et al. (1986). One of the conditional problems was arbitrary (for one of the forms, the Wason card selection problem was used), one was couched in language that was expected to encourage causal reasoning, and one was couched in language that was expected to encourage application of the permission schema. An example of a permission-schema version of the conditional problem is presented in Appendix A. The biconditional problem was an assessment of the necessary-and-sufficient checking procedure.

Verbal reasoning. A fourth set of questions was intended to measure general verbal abilities to recognize arguments, evaluate evidence, and detect analogies. These items were similar to GRE verbal exam items taken by psychology and chemistry students, and to verbal exam items on the LSAT (taken by law students) and the MCAT (taken by medical students). Many items were in fact drawn from practice materials for these verbal exams. An example of a verbal reasoning question is presented in Appendix A.

Scale Construction

Seven items were answered correctly by 90% or more of first-year students in one or more of the disciplines. In order to prevent ceiling-effect problems, all such items were dropped. One additional item was dropped because it correlated negatively with total score, that is, the sum of all items on the test minus the item itself.

Individual statistical reasoning items, as it turned

out, were no more highly correlated with the sum of other statistical items than with the sum of methodological reasoning items, nor were individual methodological items correlated more highly with the sum of other methodological items than with the sum of statistical items. These results, together with the fact that the two scales behaved very similarly for each discipline for both the cross-sectional and longitudinal designs, led us to collapse them into a single scale. All but two items on this scale were correlated more highly with the total statistical–methodological reasoning score than with either the conditional reasoning scale or the verbal reasoning scale.³ All the conditional reasoning items correlated more highly with the conditional reasoning total (minus each item itself) than with either the statistical–methodological total or the verbal total.

Study 2: Longitudinal Design

A second study was identical in every respect to the first except that it had a longitudinal design. First-year students were retested at the beginning of their third year of study. All first-year psychology students and chemistry students from the cross-sectional study were asked to be retested. Stratified random samples were drawn of the first-year medical and law students originally tested, with the stratification being based on total scores on the initial test. Only the sampled students were contacted and asked to be retested. Students initially given Form 1 were given Form 2 on the retest and vice versa. Response rates were as follows: 77 of the 88 randomly selected law students (88%), 87 of the 91 randomly selected medical students (96%), 24 of the 25 initially tested psychology students (96%), and 18 of the 31 initially tested chemistry students (58%, but 100% of the 18 still enrolled).

Students were tested under the same circumstances as third-year students in the cross-sectional study and were paid \$20 for their participation.⁴

Normative Questions About Statistical–Methodological Items

We have sometimes been asked how it is we know that the answers we endorse for the statistical and methodological questions are the “correct” ones. It would be

possible for us to say simply that we are reporting changes and that others may decide whether these changes are for the better. Undoubtedly, others will decide for themselves, regardless of what we say, but we have a normative stance, which applies to the verbal and conditional questions as well, that readers may wish to consider.

Generally speaking, the position that testers take on the question of how to identify “right answers” is that an answer endorsed by a consensus of experts is the correct one. In some cases, this presumption is based on actual polling of experts, but in most cases it is based on an assumed ability to mimic the stance of experts toward test items. We subscribe to this position.

In the present case, we have discussed many of our statistical and methodological items with experts in statistical and methodological reasoning, in several instances at conferences of experts in inductive reasoning, including statisticians as well as epistemologists and cognitive psychologists with substantial statistical expertise. For each of these items, the overwhelming consensus of the experts is that the answer we prefer is the correct one. The other items we used resemble those items to a considerable degree.

Our judgments in these respects are supported by some empirical data. First, it turns out that people who endorse the statistical answers we prefer for questions of the present type are more likely to endorse parallel answers for quite noncontroversial statistical questions involving the behavior of randomizing devices (Jepson, Krantz, & Nisbett, 1983). In addition, people who endorse the answers we prefer receive generally higher scores on intelligence tests than do those who do not. Indeed, a subtest of statistical items correlates about as well with combined verbal and mathematical ability as measured by the SAT as do some standard IQ subtests such as spatial reasoning and analogies (Jepson et al., 1983). In the present study, all of our subtest scores were positively correlated both with full test scores and with the admissions test score, which was the total GRE score (math plus verbal plus analytic test scores) in the case of psychology and chemistry students, the total MCAT in the case of medical students, and the LSAT in the case of law students. Thus, our preferences are in line both with those of experts and with those students in each discipline scoring highest on traditional ability tests.

Results

Cohort Differences

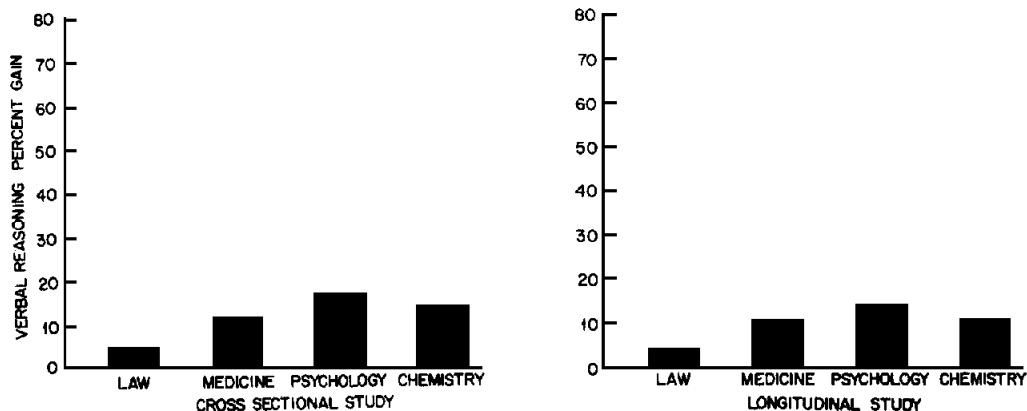
In order to determine whether there were any cohort differences in general ability, we compared the admission test scores of first-year students and third-year students within each discipline. The combined GRE score was used for psychology and chemistry students, the combined MCAT for medical students, and the LSAT for law students. There were no differences between first- and third-year students of even marginal significance for any discipline.

³ Of the two exceptions, one correlated .12 with the statistical–methodological total (minus itself) and .18 with the conditional total, and the other correlated .09 with the statistical–methodological total and .12 with the conditional total. Because the manifest content of these items was so similar to the other statistical–methodological items, we decided to treat these small discrepancies as measurement error and to leave the items on the statistical–methodological scale.

⁴ A separate study was done of the effect of retesting on scores on each of the subscales. The study included 129 medical students, none of whom had participated in either Study 1 or Study 2. The two test administrations were approximately six weeks apart. The group posttest score was not higher than the group pretest score for any of the three reasoning subtests of interest. A similar “retest” study was done with approximately 80 law students with identical results. These findings mean that any changes registered for the longitudinal study are due to something associated with two years of education in the respective disciplines rather than with the effect of retesting per se.

Figure 1

Change Scores for Verbal Reasoning as a Function of Design, Year, and Graduate Program



Initial Differences

The first-year students differed marginally by discipline in their initial scores on the conditional reasoning test, $F(3, 395) = 2.2, p = .08$.⁵ This was due to the law students scoring slightly higher and the medical students slightly lower than the other two groups. The first-year students differed substantially in both statistical-methodological and verbal reasoning, $F(3, 395) = 6.9, p < .0002, F(3, 395) = 2.93, p < .03$, respectively. These latter two differences were due almost entirely to the relatively low scores of chemistry students (though there was also a slight elevation of law student scores over the scores of others on verbal reasoning). As might be expected, it was disproportionately the low-scoring chemistry students who subsequently left the program. The combined GRE scores for chemistry students who left the program ($M = 1,676$) was also lower than for those who remained ($M = 1,880$), $t(40) = 39.18, p < .01$. As a consequence, when only those first-year students who were included in the longitudinal sample were examined, a different pattern emerged. There were no disciplinary differences in the first-year scores of those in the longitudinal sample in either statistical-methodological reasoning, $F(3, 205) = 1.78, p = .15$ or conditional reasoning, $F(3, 205) = .31, p = .82$. There remained a slight difference in verbal reasoning for this group, $F(3, 205) = 3.04, p < .05$, that was due entirely to the law students having higher scores than psychology, chemistry, or medical students.

This pattern of first-year differences means that the cross-sectional results slightly overestimate improvement for the (self-selected) third-year chemistry students, at least for statistical-methodological reasoning and verbal reasoning. Other than that, the pattern indicates that there was little difference among first-year students in either

design except for a slight advantage for the first-year law students in verbal reasoning.

Changes in Reasoning Scores

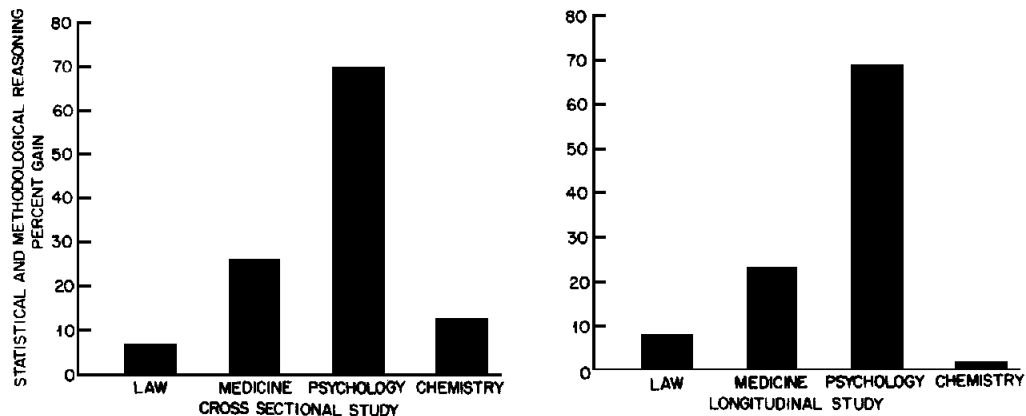
Verbal reasoning. The change scores for verbal reasoning for all students, for both the cross-sectional study and the longitudinal study, are presented in Figure 1. It may be seen that, as anticipated, all groups show slight improvement in verbal reasoning, ranging from 5% to 17% in the cross-sectional design and from 4% to 14% in the longitudinal design. Only the medical students' improvement reaches statistical significance as measured by t test, $p < .05$, for both the cross-sectional and the longitudinal study. The interaction between discipline and year fell far short of significance. These results provide a base from which to examine the differential change for the four groups in the other two types of reasoning.

Statistical-methodological reasoning. It will be recalled that statistical and methodological reasoning scores could not be differentiated from one another psychometrically. The two types of scores were also affected in the same way by training in the different disciplines. For these reasons, the scores for statistical and methodological reasoning were combined for purposes of analysis. Change scores for all subjects, for both the cross-sectional study and the longitudinal study, are presented in Figure 2. It may be seen that the psychology students changed dramatically from the first year to the third year in statistical-methodological reasoning, $t(56) = 5.00, p < .0001$ in the cross-sectional study, paired- $t(23) = 5.37, p < .0001$ in the longitudinal study. Medical students also changed substantially, $t(173) = 2.73, p < .01$ in the cross-sectional study, paired- $t(86) = 3.09, p < .005$ in the longitudinal study. Law students, in contrast, showed no significant improvement in statistical-methodological reasoning, $t(261) < 1$ in the cross-sectional study, paired- $t(76) = 1.17, ns$, in the longitudinal study. Chemistry students also did not improve, t s for both designs < 1 . The interaction between year and discipline is significant for both

⁵ All p values reported are based on two-tailed tests. Means (and more detailed summaries) of the data are available upon request from the first author.

Figure 2

Change Scores for Statistical and Methodological Reasoning as a Function of Design, Year, and Graduate Program



the cross-sectional design, $F(3, 552) = 5.33, p < .001$, and the longitudinal design, $F(3, 202) = 5.62, p < .001$.

Conditional reasoning. Figure 3 presents change in conditional reasoning for all students, for both the cross-sectional and the longitudinal design. The students in law, medicine, and psychology all improved in their ability to reason about conditional problems. None of the differences are significant in the cross-sectional study, but all three are significant in the longitudinal study, paired- $t(76) = 2.91, p < .005$ for law, paired- $t(86) = 3.22, p < .002$ for medicine, paired- $t(23) = 2.29, p < .05$ for psychology. It should also be noted that the trend was for all three groups to improve on all four types of questions—arbitrary, causal wording, permission wording, and bi-conditional. In contrast, chemistry students did not improve in conditional reasoning, both t s < 1 . The inter-

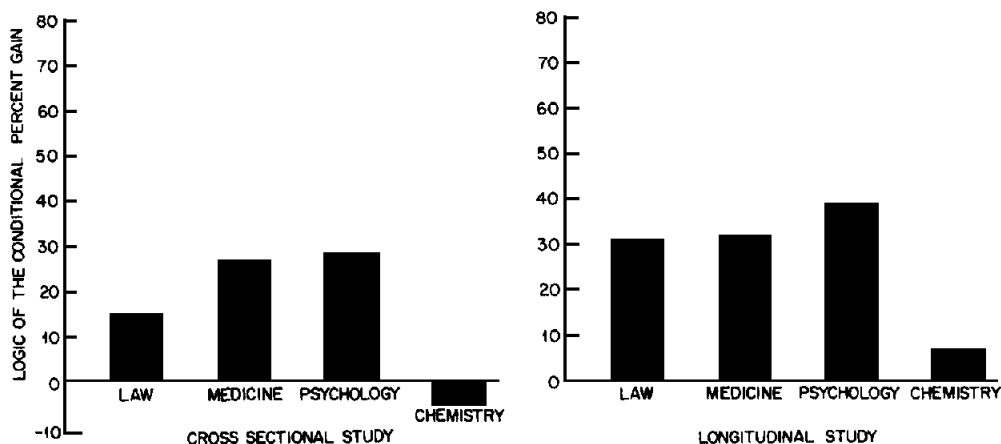
action between year and discipline fell short of significance for both the cross-sectional and the longitudinal design.

Discussion of Studies 1 and 2

The results are thus quite consistent with the view that reasoning can be taught and that different graduate disciplines teach different kinds of reasoning to different degrees. It appears that the probabilistic sciences of psychology and medicine teach their students to apply statistical and methodological rules to both scientific and everyday-life problems, whereas the nonprobabilistic science of chemistry and the nonscientific discipline of the law do not affect their students in these respects. Psychology, medicine, and the law all seem to teach their students rule systems that increase their ability to reason about problems for which the conditional or the bicon-

Figure 3

Change Scores for Conditional Reasoning as a Function of Design, Year, and Graduate Program



ditional in logic affords solutions. In our view, this is not because these fields teach the rules of formal logic, but because they teach pragmatic reasoning schemas that provide the same solutions as does the conditional. These reasoning schemas include causal schemas and their associated evidence-checking procedures, and contractual schemas, including the permission and obligation schemas.

The results for the cross-sectional and longitudinal designs are similar enough to give us some confidence that the results are an accurate reflection of true changes produced by training in the various disciplines at the University of Michigan. It would increase our confidence in the generality of the results, however, to have a replication at another institution. This is especially the case for the chemistry results, which amount to a null finding on a relatively small *N* base. In order to increase our confidence in the findings and to ensure that no features peculiar to a Michigan education were important, the study was partially replicated at the University of California at Los Angeles (UCLA).

Cross-Sectional Replication: Study 3

Method

The method for UCLA was identical in every respect to that for the cross-sectional study at the University of Michigan, except that only psychology and chemistry students were studied and the instrument was shortened to delete items having ceiling effects at Michigan. In addition, an extra biconditional item was added to the conditional-reasoning scale on both forms. A very high response rate was obtained. The *n*s were 27 of 30 first-year psychology students (90%), and 27 of 27 still-enrolled third-year psychology students (100%), 35 of 37 first-year chemistry students (95%), and 28 of 30 still-enrolled third-year chemistry students (93%). The drop-out rates were lower for chemistry at UCLA (22%) than at Michigan (46%). The psychology drop-out rate at UCLA (23%) was virtually identical to that for chemistry.

Results

Cohort differences. A comparison of combined GRE scores for first- and third-year students in the two disciplines showed only very slight differences between chemistry cohorts but a nontrivial difference between psy-

chology cohorts, $t(47) = -1.91, p = .06$, with first-year students ($M = 2,002$) scoring higher than third-year students ($M = 1902$). Because GRE scores are correlated with each of the dependent variables, results were subjected to covariance analyses and adjusted means are reported. GRE scores were not available for 12 students, therefore *ns* for the covariance analyses are slightly reduced.

Social science versus natural science differences among psychology students. Preliminary analyses revealed that the magnitudes of training effects for psychology students were smaller at UCLA than at Michigan. Subsequent analyses indicated that differences were entirely due to the relatively large fraction of "natural science" psychology students at UCLA. Fully 35% of UCLA psychology students were in the areas of physiological, learning and behavior, or experimental psychology, versus only 12% at Michigan. Analyses revealed that these fields showed little positive change in reasoning scores, whereas the "social science" fields of personality, social, developmental, and clinical showed changes fully comparable to those observed at Michigan. Results are therefore presented separately for social science and natural science psychologists. The *n* for first-year social science psychology students was 17, for third-year social science psychology students 15, for first-year natural science psychology students 8, for third-year natural science psychology students 9, for first-year chemistry students 31, and for third-year chemistry students 25.

Verbal reasoning. Table 1 presents reasoning scores, reported as percent correct for each scale, for both social science and natural science psychology students, and for chemistry students. It may be seen that the pattern for verbal reasoning is very similar to that observed in Studies 1 and 2, that is, all fields of study show slight and non-significant improvements from the first year to the third year, all *ts* < 1. The interaction between year and discipline fell far short of significance.

Statistical and methodological reasoning. It may be seen in Table 1 that the statistical-methodological reasoning of social science psychologists showed substantial change, amounting to a 59% improvement over two years of study, $F(1, 31) = 11.35, p < .002$. In contrast, neither the natural science psychologists nor the chemistry students improved. The natural science psychologists actually showed a slight but nonsignificant decrease. The interaction between year and discipline is highly signifi-

Table 1
Adjusted Means for Reasoning Scores for First- and Third-Year UCLA Psychology and Chemistry Students

Type of reasoning	Social science psychologists		Natural science psychologists		Chemistry	
	1st year	3rd year	1st year	3rd year	1st year	3rd year
Verbal	51.2	55.6	42.4	45.7	50.4	51.1
Statistical and methodological	40.5	64.2	47.4	36.1	29.0	33.8
Conditional	32.7	51.6	53.3	36.1	39.3	31.3

cant, $F(2, 104) = 5.61, p < .005$. It should be noted that the difference between social science and natural science psychologists is particularly marked for the questions having everyday-life content, interaction $F(1, 48) = 9.15, p < .005$, and is substantially less for questions having scientific content, interaction $F(1, 48) = 2.84, p = .10$.

Conditional reasoning. It may be seen in Table 1 that conditional reasoning for the social science psychologists improved substantially over two years of study, $F(1, 31) = 3.93, p = .06$. Conditional reasoning for both natural science psychologists and chemistry students actually declined, although the decline is not significant in either case. The interaction between year and discipline is significant, $F(2, 104) = 3.36, p < .05$.

Discussion of Study 3

The results for the replication study lend support to the results, and the interpretation of the results, for Studies 1 and 2. Change in verbal reasoning is positive and slight for all fields. Change in both statistical-methodological reasoning and conditional reasoning is positive and pronounced for social science psychologists and nil or slightly negative for chemistry students and natural science psychologists. The results indicate that the early years of training in the probabilistic sciences can be helpful for reasoning about certain kinds of problems using particular rule systems whereas the early years of natural science training have little effect on those same problems and rule systems.

General Discussion

The results show that training in both psychology and medicine can affect statistical reasoning about everyday-life problems, methodological reasoning about everyday-life problems employing a number of rules related to the confounding principle, and reasoning about problems that logicians can solve using the material conditional in deductive logic. In addition, training in the law affects conditional reasoning. Training in chemistry does not seem to affect any of these kinds of reasoning.

Why do we find this pattern of results? Psychologists receive statistical training as part of the first two years' research experience; thus, the finding that training in psychology affects statistical reasoning about everyday-life events amounts to an extension of previous results by Fong et al. (1986) showing that statistical training affects thinking about everyday-life events. Most psychologists will not be surprised that training in their field also affects reasoning about confounded variables in everyday life involving methodological principles such as selection and the need for control groups. Particularly in the social science branches of psychology, day-to-day research experience requires students to think about everyday-life events of the kind being studied and to employ such methodological principles. Although the natural science branches of the discipline also teach these rules in a formal way, we believe they provide much less experience in applying them to everyday-life events. Thus, it would not be surprising if further work replicated the results of Study

3, which showed more improvement for social science psychology students than for natural science psychology students, especially in working with everyday-life problems. (For the time being, however, the reader should be warned that this difference was found only at one institution and that the n of natural science psychology students was rather small.)

It may be more surprising to psychologists that medical training increases statistical and methodological reasoning skills as much as it does. Examination of the medical curriculum at the University of Michigan reveals, however, that a great deal of training is geared toward teaching students how to think about uncertain events. For one thing, students are given a pamphlet explaining the basics of statistics and are expected to learn its contents. In addition, they are given original articles to read, with their reports of statistical tests, considerations of variance, sample size, and so forth. Perhaps more important, although most of the course content in the first two years is ostensibly focused on the natural sciences, there is in fact a continual effort to relate such content to complicated everyday medical problems. This attempt contains frequent allusions to the difficulties of making inferences in medical situations and to the uncertainties inherent in medical practice. Students are told of false positive and false negative rates for tests and of costs for tests, and they are asked whether, for a patient with particular presenting symptoms, they would "pay" for a particular test to be done. It is also made clear to students that the confounded variable principle underlies much medical "sleuthing." For example, they are reminded that some behaviors, such as cigarette smoking, are likely to be associated with others, such as coffee drinking. Thus, it is not surprising that two years of medical school affects statistical and methodological reasoning about everyday-life events. Because such training is also likely to require students to think about causal relations of different kinds, including those for which the evidence-checking procedures resemble those for the conditional and those for the biconditional, it is also not surprising that medical training results in improved ability to solve conditional and biconditional problems. Probably for the same reason, training in psychology results in improved ability to solve these problems.

Training in the law does not stress rules for dealing with variability or uncertainty in causal relations, and so it is not surprising that it produces no improvement in the ability to apply the statistical and methodological rules of the probabilistic sciences to either scientific studies or everyday-life events. Legal training does provide substantial instruction and drill in the logic of permissions and obligations, which can be used to solve problems in the conditional, and it provides additional instruction and drill for other contractual relations, particularly those in which an action must be taken if and only if some other event occurs, which can be used to solve biconditional problems.

In our view, training in chemistry provides no improvement in statistical or methodological reasoning be-

cause training in that field, especially in the early years, does not deal with events that are probabilistic. Statistics is not taught in the first two years of chemistry, and one would be hard pressed to come up with examples of confounded-variable rules in the curriculum. Similarly, there is little need to differentiate among the various types of causal relations because chemistry deals primarily with necessary-and-sufficient causes. Moreover, to the extent that any statistical or conditional reasoning does occur, it would be in field-specific settings highly abstracted from the problems of everyday life. The chemistry curriculum obviously should not be regarded as deficient, any more than a failure to teach about electrical charges should be regarded as a deficiency of psychological or legal training. However, the luxury of not being confronted with messy problems that contain substantial uncertainty and a tangled web of causes means that chemistry does not teach some rules that are relevant to everyday life.

The present results indicate that 20th century psychologists have been too quick to conclude that formal discipline is not possible and that rule training has little generalized educational potential. The results make it clear that training of some kinds has substantial effects on the way people reason about some sorts of problems. The error of psychologists in this century, in our view, has been to assume that because some kinds of rule systems do not generalize readily from problems having certain features to problems having another set of features (for example, rules for estimating the area of rectangles or rules for solving the Tower of Hanoi problem), *all* rules show equally poor transfer and generalization properties. The failure, in other words, amounts to a willingness to endorse the null hypothesis on the basis of a limited amount of research on a relatively small set of rules. Importantly, none of the studies that have led to the pessimistic perspective on the value of formal discipline have examined situations in which people learn through immersion in a field of study and have numerous occasions to apply the rules of the discipline to problems that arise both inside and outside their course of study. Yet this seems to be how disciplinary learning naturally occurs.

What types of inferential rules can be taught in such a way that they will be graceful additions to the rule systems that people already use? We think that a major class of such rules are those that people have induced, though only partially, in the course of their daily existence. Rules about assessing causality, rules for generalizing, rules for determining argument validity, and rules for assessing the probativeness of evidence are the kinds of rules that people must have in some measure in order to live effectively in the world.

Our work predisposes us to be optimistic about the possibility of identifying abstract rule systems underlying solutions of real-world problems and about the ability of educators to teach these rule systems. This optimism parallels that stemming from recent work by Herrnstein, Nickerson, de Sanchez, and Swets (1986) and others (see Nickerson, Perkins, & Smith, 1985, for a review) who have shown that strategies of reasoning can be taught to

elementary school children. These investigators have taught such foundations of reasoning as how to use dimensions to analyze and organize similarities and differences, how to recognize and extrapolate different types of sequences, how to see the structure of simple propositions and analyze complex arguments, how to evaluate consistency, and so on. Such training not only improves performance on IQ tests but also improves the quality of open-ended written arguments.

The truth is that we know very little about reasoning and how to teach it. The one thing we thought we knew—namely, that formal discipline is an illusion—seems clearly wrong. Just how wrong, and therefore just how much we can improve reasoning by instruction, is now a completely open question.

REFERENCES

- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391–416.
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, *18*, 293–328.
- Cheng, P. W., Nisbett, R. E., & Oliver, L. M. (1987). *The use of causal schemas in the evaluation of evidence for hypotheses*. Unpublished manuscript, University of California, Los Angeles.
- Costa, R. M. (1982, March 6). Latin and Greek are good for you. *The New York Times*, p. 23.
- D'Andrade, R. (1982). Paper presented at the Symposium on the Ecology of Cognition: Biological, Cultural, and Historical Perspectives, Greensboro, NC.
- Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*, 253–292.
- Fong, G. T., & Nisbett, R. E. (1988). *Domain specificity and domain independence of statistical training: The effects of time*. Unpublished manuscript, Northwestern University.
- Griggs, R. A., & Cox, U. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, *73*, 407–420.
- Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (pp. 21–41). Hillsdale, NJ: Erlbaum.
- Herrnstein, R. J., Nickerson, R. S., de Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist*, *41*, 1279–1289.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. T. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: Bradford Books/MIT Press.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books. (Original work published 1955)
- Jepson, C., Krantz, D. H., & Nisbett, R. E. (1983). Inductive reasoning: Competence or skill? *Behavioral and Brain Sciences*, *6*, 94–101.
- Jowett, B. (1937). *The dialogues of Plato*. New York: Random House.
- Kelley, H. H. (1971). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, *28*, 107–128.
- Manktelow, K. I., & Evans, J. St. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, *63*, 395–400.
- Mann, L. (1979). *On the trail of process*. New York: Grune & Stratton.
- Nickerson, R. S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.
- Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, *238*, 625–631.
- Nisbett, R. E., Krantz, D. H., Jepson, D., & Kunda, Z. (1983). The use

of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339–363.

Reed, S. K., Dempster, A., & Erttinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 106–125.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology*, 6, 436–450.

Share, K. W. (1979). The primary mental abilities in adulthood: An exploration in the development of psychometric intelligence. *Life Span Development and Behavior*, 2, 67–115.

Thorndike, E. L. (1906). *Principles of teaching*. New York: A. G. Seiler.

Thorndike, E. L. (1913). *The psychology of learning*. New York: Mason-Henry.

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

Appendix A

Examples of Items Used

(Correct answers are denoted by †)

Statistical Reasoning—Everyday Life

After the first two weeks of the major league baseball season, newspapers begin to print the top ten batting averages. Typically, after two weeks, the leading batter has an average of about .450. Yet no batter in major league history has ever averaged .450 at the end of a season. Why do you think this is?

†(a) A player's high average at the beginning of the season may be just a lucky fluke.

(b) A batter who has such a hot streak at the beginning of the season is under a lot of stress to maintain his performance record. Such stress adversely affects his playing.

(c) Pitchers tend to get better over the course of the season, as they get more in shape. As pitchers improve, they are more likely to strike out batters, so batters' averages go down.

(d) When a batter is known to be hitting for a high average, pitchers bear down more when they pitch to him.

(e) When a batter is known to be hitting for a high average, he stops getting good pitches to hit. Instead, pitchers "play the corners" of the plate because they don't mind walking him.

Methodological Reasoning—Everyday Life

The city of Middleopolis has had an unpopular police chief for a year and a half. He is a political appointee who is a crony of the mayor, and he had little previous experience in police administration when he was appointed. The mayor has recently defended the chief in public, announcing that in the time since he took office, crime rates decreased by 12%. Which of the following pieces of evidence would most deflate the mayor's claim that his chief is competent?

†(a) The crime rates of the two cities closest to Middleopolis in location and size have decreased by 18% in the same period.

(b) An independent survey of the citizens of Middleopolis shows that 40% more crime is reported by respondents in the survey than is reported in police records.

(c) Common sense indicates that there is little a po-

lice chief can do to lower crime rates. These are for the most part due to social and economic conditions beyond the control of officials.

(d) The police chief has been discovered to have business contacts with people who are known to be involved in organized crime.

Conditional Reasoning—Permission Schema

You are a public health official at the international airport in Manila, capital of the Philippines. Part of your duty is to check that every arriving passenger who wishes to enter the country (rather than just change planes at the airport) has had an inoculation against cholera. Every passenger carries a health form. One side of the form indicates whether the passenger is entering or in transit, and the other side of the form lists the inoculations he or she has had in the past six months. Which of the following forms would you need to turn over to check? Indicate only those forms you would have to check to be sure.

Transit	Entering	Inoculated against: cholera hepatitis	Inoculated against: typhoid
---------	----------	--	-----------------------------------

Box 1

Box 2

Box 3

Box 4

(a) Boxes 2 & 3

(b) Box 2 only

(c) Boxes 2, 3, & 4

†(d) Boxes 2 & 4

(e) Box 3 only

Verbal Reasoning

The new miracle drug Amotril has caused unforeseen side effects of a devastating nature; therefore, no new drugs should be released for public consumption without a thorough study of their effects.

Which of the following arguments most closely resembles the argument above?

(a) Because exposure to several hours of television a day has been shown to undermine children's interest in reading, children should be prevented from watching television.

(b) Because it is difficult to predict whether the results of pure research will be of practical benefit to human beings, the amount of money spent on such research should be sharply curtailed.

(c) The 1977 model of this compact station wagon has been shown to have a faulty exhaust system; therefore, it is urgent that this model be recalled immediately.

†(d) Some of the worst highway accidents have been caused by teenagers between the ages of 16 and 18; therefore, only carefully screened members of this age group should be granted driver's licenses.

(e) Rising medical costs have put many routine medical procedures out of the reach of low- and middle-income families; therefore, doctors should prescribe only the most essential laboratory tests.