A Longitudinal Study of the Effects of Undergraduate Training on Reasoning

Darrin R. Lehman
University of British Columbia

Richard E. Nisbett
University of Michigan

The effects of undergraduate training in the natural sciences, humanities, and social sciences on inductive reasoning requiring the use of statistical and methodological principles and on reasoning about problems in conditional logic were examined. Social science training produced large effects on statistical and methodological reasoning, whereas natural science and humanities training produced smaller, but still marginally significant, effects. Natural science and humanities training produced large effects on ability to reason about problems in conditional logic, whereas social science training did not. The improvement in conditional reasoning among natural science students appears to be due, in large part, to knowledge gained in mathematics courses. The results indicate that inferential rule systems, as taught by various fields, can affect reasoning about a wide range of problems. The findings lend support to a version of the premodern *formal discipline* hypothesis.

Does an undergraduate education improve reasoning about everyday-life problems? Do some forms of undergraduate training enhance certain types of reasoning more than others? These issues have not been addressed in a methodologically rigorous manner (Nickerson, Perkins, & Smith, 1985). We consequently have little knowledge of whether reasoning can be improved by instruction, yet this question has long been regarded as central to theories of cognitive development.

Formal Discipline and Opposing Views

Until the 20th century, most educators endorsed the concept of *formal discipline,* initially formulated by Plato. This is the notion that reasoning can be improved as a result of teaching rules within a particular field, which then generalize outside the bounds of that field. This view is consistent with the optimistic hypothesis that instruction in abstract rule systems can affect reasoning about everyday-life events (Lehman, Lempert, & Nisbett, 1988; see also Fong, Krantz, & Nisbett, 1986; for a general summary, see Nisbett, Fong, Lehman, & Cheng, 1987). The formal discipline concept was countered initially by Thorndike (Thorndike, 1906, 1913; Thorndike & Woodworth, 1901), who concluded that people use only highly domain-specific rules dealing with concrete types of events to solve problems. Many current theorists endorse Thorndike's view. For example, Newell (1980) holds that "the modern . . . position is that learned problem-solving skills are, in general, idiosyncratic to the task." Piaget (see, for example, Inhelder & Piaget, 1955/1958) put forth a highly influential theory suggesting that

people do use abstract inferential rules, but because learning of these rules comes mainly through self-discovery (constrained by level of cognitive development), formal training can do little to extend it. Thus, much contemporary theorizing has suggested that reasoning is either domain-specific or quite difficult to teach. Such theorizing gives us little reason to expect that higher education would have much effect on reasoning abilities.

Our previous results (Lehman et al., 1988), however, suggest that a version of the formal discipline view is correct. We reported on the effects of graduate training in law (a nonscience), medicine and psychology (two probabilistic sciences), and chemistry (a nonprobabilistic or deterministic science) on statistical and methodological reasoning and on reasoning about problems in the logic of the conditional. We found that graduate students in the probabilistic sciences improved in statistical and methodological reasoning and that students in both the probabilistic sciences and in the law improved in conditional reasoning. Chemistry students did not improve in any of the types of reasoning examined.

Statistical and Methodological Reasoning

Statistical reasoning includes a family of related schemas or heuristics that are derivable from the law of large numbers. These include the rule that, all things being equal, larger samples are more representative of the population they are drawn from than are smaller samples and that this difference is greater the greater the variability inherent in the population. The related rule of regression to the mean and the base-rate principle are derivable directly from the law of large numbers and the concept of variability (Nisbett, Krantz, Jepson, & Kunda, 1983). For example, regression to the mean occurs because extreme values for an object or sample are due partly to observational error and therefore are less likely to be extreme when the object is reexamined or a new sample is observed.

The sort of methodological reasoning examined in our previous work included judgments requiring an understanding of problems posed by confounded variables. These reasoning

skills are particularly helpful when thinking about causes that are neither necessary nor sufficient (Kelley, 1971, 1973). They include rules about when and how to use control groups, how to avoid sample bias, and how to recognize and avoid 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 (third-variable problems).

Conditional Reasoning

Conditional reasoning, or reasoning using the material conditional in formal logic, is a form of deductive logic based on modus ponens and the contrapositive or modus tollens rule. Wason (1966) established that people typically have great difficulty with abstract problems that follow the form of the material conditional "if p, then q." He examined how people respond to abstract selection tasks that embody this logic. For example, subjects are shown four cards displaying A, B, 4, and 7; are told that all cards have letters on the front and numbers on the back; and are asked to turn over as many cards 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 was not an even number on the back, the rule would be violated) and the 7 (because if there was 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).

Recently, Cheng, Holyoak, and their colleagues (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986) argued that, instead of using the rules of formal logic to solve conditional problems, people use pragmatic inferential rules, or pragmatic reasoning schemas, which capture recurring regularities among problem goals and among event relations that people encounter in everyday life. These rules are fully abstract in that they are not tied to any content domain, but they are not as independent of relation types and problem goals as formal logical rules or the purely syntactic rule systems often studied by modern cognitive psychologists.

One type of pragmatic inferential rule system is the *contractual schema*. Contractual schemas include rules about 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 map onto the conditional in formal logic. That is, 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" holds. Subjects can solve problems invoking the obligation schema that map directly onto the conditional. For example, most subjects know how to check whether the rule "if X occurs, then you must do Y" has been violated. In addition, research (Cheng & Holyoak, 1985; Cheng et al., 1986) has shown that whereas training in formal logic has little effect on people's ability to solve either arbitrary or semantically meaningful versions of

conditional problems, training in pragmatic reasoning schemas has a substantial effect.

Cheng, Morris, and Nisbett (1990) 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 necessity 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. It is interesting that one of the schemas, namely, that for sufficient-but-not-necessary causes, has checking procedures identical to those for the conditional selection task ("if p, then q"). A hypothesis of the sufficient-but-not-necessary type, "Factor A always (though not exclusively) causes Effect B," can be disproven by examining A and finding that not-B was the case or by examining not-B and finding that A was the case. Cheng et al. (1986) discovered 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, it turns out, 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 conditional reasoning materials used in this study include (a) classic, arbitrary material conditional problems of the Wason type; (b) biconditional problems; and conditional problems having (c) permission or (d) causal interpretations.

Effects of Graduate Training

The results of Lehman et al. (1988) showed improvement on different types of reasoning tasks by graduate students, in terms of rule systems taught by the various fields. Lehman et al. argued that psychology and medicine improve statistical and methodological reasoning about a wide range of problems because both fields teach their students how to think about uncertain events, in part through instruction in statistics. In addition, in the case of psychology, students are instructed in methodology and given day-to-day hands-on research experience. In the case of medicine, a concerted effort is made to have students read research articles and understand the complicated statistical and methodological connections to everyday medical problems. Because neither training in law nor chemistry stresses rules for dealing with variability or with uncertainty in causal relations, it was not surprising that these two disciplines did not improve students' abilities to apply statistical or methodological rules to the range of problems.

Lehman et al. (1988) argued that students of psychology and medicine improved on conditional reasoning because their graduate training in the probabilistic sciences required them to think about causal relations of different kinds, including those for which the evidence-checking procedures resemble those for the conditional. They argued that law students improved in conditional reasoning because law provides substantial instruction in thinking about contractual relations that have the form of the conditional, such as permissions and obligations. Finally, Lehman et al. argued that chemistry students did not

improve in conditional reasoning because chemistry provides neither contractual training nor training in how to reason about the wide variety of types of causal relations encountered in the probabilistic sciences.

Implications for the Effects of Undergraduate Training

In this undergraduate study, we again wanted to include probabilistic sciences, nonprobabilistic (or deterministic) sciences, and nonsciences. Social science training, in addition to psychology, included anthropology, economics, political science, and sociology. Natural science training included biology, chemistry, microbiology, and physics. Humanities training included communications, English, history, journalism, linguistics, and philosophy. We used a longitudinal design, following subjects through 4 years of undergraduate training. This allowed us to remove the possibility of alternative explanations such as differential attrition rates.

We anticipated that psychology, and, more generally, social science training, at the undergraduate level would parallel the relative improvement in statistical and methodological reasoning of psychology training at the graduate level. We anticipated that training of this sort would alert people to consider sample sizes when making inferences and would also affect methodological reasoning.

In contrast, we anticipated that undergraduate training in the deterministic sciences and in the nonsciences would have less effect on statistical and methodological reasoning. The natural sciences focus primarily on deterministic causes and consequently should produce less improvement in rule usage concerned with uncertainty and self-selection. The humanities, with their emphasis on language, history, and philosophy, should also produce less improvement on these statistical and methodological reasoning abilities.

Lehman et al. (1988) discovered improvements in conditional reasoning among graduate students in the probabilistic sciences and attributed these improvements to extensive experience with causal relations of various kinds (e.g., sufficient but not necessary causes). Cheng et al. (1990) have recently found that psychology graduate students do indeed improve in their ability to reason causally, whereas chemistry and philosophy graduate students do not. Should we expect the same sort of improvement in conditional reasoning for social science majors at the undergraduate level? It seems unlikely that the training in causal reasoning at the undergraduate level is nearly as intensive as it is at the graduate level. For example, undergraduate training in psychology usually does not involve large doses of experimental research. Thus, although graduate students in the probabilistic sciences improve in conditional reasoning (presumably because of their improved ability to reason about formally different types of causal relations), it is not clear that similar improvements should occur for social science undergraduates.

What about predictions concerning improvements in conditional reasoning for natural science and humanities undergraduates? Lehman et al. found no improvement in conditional reasoning for chemistry graduate students. Similarly, Cheng et al. (1990) discovered that graduate training in chemistry did not improve students' abilities to reason causally. On the other hand, recent research has suggested a specific prediction concerning improvements in conditional reasoning that probably applies to undergraduates in the natural sciences. Jackson and Griggs (1988) found a significant effect of mathematics training on ability to solve conditional problems. They found that subjects trained in mathematics performed better than subjects trained in computer science, electrical engineering, and the social sciences. Jackson and Griggs reasoned that mathematics students, with their experience in proof by contradiction ("if p were the case, q would have to be the case; not q, therefore not p"), are quite familiar with the basic material conditional relation. In effect, Jackson and Griggs argued, proof by contradiction is a pragmatic reasoning schema for students with training in math. Given that the conditional task maps onto this schema, it is possible that mathematics students would perform at a higher level on conditional problems. Therefore, natural science undergraduates, because of the greater number of mathematics courses in their curriculum, may improve significantly in conditional reasoning. Moreover, it should follow that the more math courses natural science majors take, the more they will improve in conditional reasoning. We have no reason to expect that humanities undergraduates will gain much in conditional reasoning because they are not taught extensively about causal schemas or contractual relations, and they do not take an abundance of mathematics courses.

To help test these hypotheses, we obtained the transcripts of our undergraduate subjects and recorded a number of types of courses taken by them. These included all of the courses theoretically relevant to gains in either statistical, methodological, or conditional reasoning, namely, computer science, logic, mathematics, natural science lab, psychology lab, social science lab, and statistics courses.

To ensure that improvements in statistical, methodological, and conditional reasoning by particular groups of subjects were not observed because these groups were improving in all aspects of reasoning, we examined the effects of different undergraduate training on verbal reasoning of the kind tested by the Scholastic Aptitude Test (SAT). Because verbal reasoning ability increases slightly during the young adult years for those involved in mentally challenging work (Share, 1979), we expected subjects in each of the disciplines to improve slightly, and to about the same extent, in verbal reasoning.

The above predictions regarding improvements in statistical, methodological, and conditional reasoning contradict much of the current thinking on reasoning and cognitive development. The notion that reasoning is domain-specific or that abstract rule systems cannot be taught suggests that education would have little effect on people's abilities to apply formal rules of inference to everyday-life problems. To the extent that such improvements in reasoning are discovered, a version of the pre-20th century formal-discipline theory would gain plausibility. The Piagetian view that inferential rules are learned essentially just by self-discovery would lose plausibility.

Method

Subjects

Subjects were University of Michigan undergraduates tested in the first term of their 1st year of school and again in the second term of

their 4th year of school. Subjects were contacted initially through a random selection of first-year students from a university-wide roster. Virtually all of them had recently graduated from high school. The subjects were phoned and asked if they had a tentative selection for their major of study (none of them had declared their major officially at this point). Those who reported a major that fell into one of the disciplines under investigation were then invited to participate in the study. The first-year response rates were as follows: 39 of 42 randomly sampled natural science students (93%), 46 of 50 randomly sampled humanities students (92%), 38 of 42 randomly sampled social science students (90%), and 42 of 45 randomly sampled psychology students (93%).

We attempted to recontact all of the pretest subjects approximately 3½ years later (in the middle of their senior year). Subjects were phoned, reminded about the study, and asked about their current major. If their major fell into one of the four disciplines, they were invited to participate a second time. As might be expected, by the middle of their fourth year, a number of students had switched majors and consequently were reclassified. Of the original 165 subjects who completed the pretest in their first year, 121 still had majors that we had targeted for study, and completed the posttest (73%). The sample sizes were as follows: 29 natural science students, 39 humanities students, 30 social science students, and 23 psychology students. An additional 44 subjects who took the pretest did not complete the posttest because of the following reasons: were not located (20); were repeated no-shows (11); were living out of the country (2); refused (3); and had switched to majors not included in the study (e.g., business; 8). Of the 29 natural science subjects, only 7% had switched majors since their first year. Of the 39 humanities subjects, 31% had switched majors, approximately half from psychology and half from other social sciences. Of the 30 social science subjects, 43% had switched majors, spread evenly throughout the other three categories. Finally, of the 23 psychology subjects, 22% had switched majors, approximately half from natural science and half from the humanities.

Procedure

Subjects were administered reasoning tests (one during their first year and a second during their fourth year) in a relaxed atmosphere in offices at the Institute for Social Research. They 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. Approximately half of the subjects in each group completed Form 1 of the reasoning test at the pretest, and the other half completed Form 2. Subjects initially given Form 1 were given Form 2 on the retest and vice versa.[1] The test took between 45 and 70 min to complete. Subjects were paid $5 for their participation in the pretest and $10 for their participation in the posttest.

Instrument

The reasoning test consisted of three sets of multiple-choice questions, with items spread evenly throughout the test booklet. One set was intended to measure statistical and 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 used in the SAT Verbal test.

Statistical and methodological reasoning. The statistical reasoning items were intended to measure subjects' abilities to apply the law of large numbers and the regression or base rate principle. These problems were embedded in either scientific or everyday-life contexts; scien-

tific questions were those that specifically described a study of some kind in the natural or social sciences, whereas everyday-life problems did not. Examples of a "scientific-law-of-large-numbers" problem and an "everyday-life-regression-principle" problem are presented in their entirety in the Appendix.

The methodological problems were intended to measure subjects' knowledge of selection or attrition effects, the need for control group data, problems with confounded variables, and sample bias effects. These questions were also embedded in either scientific or everyday-life contexts. Examples of a "scientific-attrition" problem and an "everyday-life-sample-bias" problem are presented in the Appendix.

Conditional reasoning. Each form contained three questions that could be solved by application of the conditional and one that could be solved by application of the biconditional (used by Cheng et al., 1986). One of the conditional problems was arbitrary (e.g., the Wason, 1966, card-selection problem), 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 "conditional-causal schema" problem is presented in the Appendix.

Verbal reasoning. A third set of questions was intended to measure general verbal abilities to recognize arguments, evaluate evidence, and detect analogies. These items were similar to SAT Verbal exam items, and some were taken from practice materials for these exams.

Scale construction. Information about the psychometric properties of the three sets of items came from the Lehman et al. (1988) study of graduate students. Individual statistical reasoning items 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. Consequently, these items were collapsed into a single scale. All but two items on this scale were correlated more highly with the total statistical–methodological reasoning score (minus each item itself) than with either the Conditional Reasoning scale or the Verbal Reasoning scale. All of the conditional reasoning items correlated more highly with the conditional reasoning total than with either the statistical–methodological total or the verbal total. In the two forms of the reasoning test, there were 11 and 9 statistical–methodological items, 4 and 4 conditional items, and 8 and 7 verbal items, respectively.

Results

Initial Differences

There were no disciplinary differences in first-year scores in verbal reasoning, $F(3, 117) = 1.04$, $p = .38$, statistical–methodological reasoning, $F(3, 117) = 0.42$, $p = .74$, or conditional reasoning, $F(3, 117) = 0.64$, $p = .59$.[2] These data were grouped on the basis of subjects' eventual majors of study.

To determine whether there were any disciplinary differences in general ability, we compared the admission test scores of subjects in the various disciplines. There were no statistically

---

[1] Separate studies were conducted of the effect of retesting on scores on each of the reasoning scales. Over 200 subjects were administered the two forms of the test approximately 6 weeks apart. The group posttest scores were not higher than the group pretest scores for any of the reasoning subsets of interest. These findings mean that any changes registered for this study were due to something associated with 4 years of undergraduate education in the respective disciplines rather than with the effect of retesting per se.

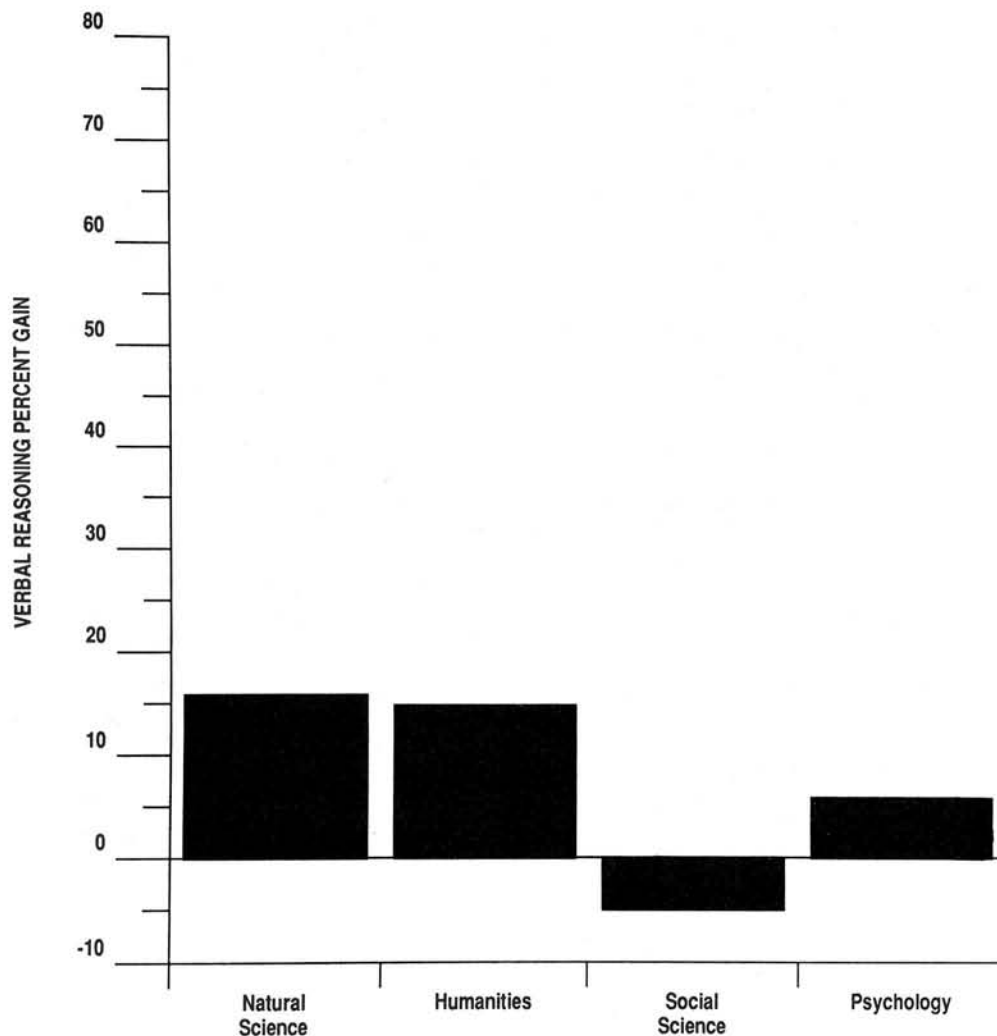[2] All p values reported are based on two-tailed tests.

Figure 1. Change scores for verbal reasoning as a function of year and undergraduate studies.

significant differences among the disciplines on either the Verbal or Quantitative sections of the Scholastic Aptitude Test. However, marginal trends were found for each of the tests ($p < .10$ for Verbal, $p < .11$ for Quantitative). The Verbal SAT score means indicate that the humanities students scored highest ($M = 582$), followed by the natural science students ($M = 560$), the social science students ($M = 540$), and the psychology students ($M = 524$). The Quantitative SAT score means indicate that the natural science students scored highest ($M = 621$), followed by the social science students ($M = 583$), the humanities students ($M = 558$), and the psychology students ($M = 553$).

Changes in Reasoning Scores

Verbal reasoning. The change scores for verbal reasoning for all groups of subjects are presented in Figure 1. It may be seen that all of the groups except social science majors showed slight improvements in verbal reasoning, ranging from 6% to 16%. The social science group actually decreased trivially (−5%).

None of the differences reached statistical significance, and the interaction between discipline and year fell far short of significance, $F(1, 117) = .82$, *ns.* These results provide a base from which to examine the differential change in the other two types of reasoning.

Statistical–methodological reasoning. The change scores for statistical–methodological reasoning for all groups of subjects are presented in Figure 2. It may be seen that the psychology and social science students changed dramatically from first year to fourth year in statistical–methodological reasoning, $t(22) = 3.96$, $p < .001$, for psychology students, and $t(29) = 4.10$, $p < .001$, for social science students. The improvement in statistical–methodological reasoning for natural science and humanities students, although much smaller than for psychology and social science students, was still marginally significant (both $ps < .10$). Pooling the probabilistic sciences (psychology and social science) and the nonprobabilistic sciences and nonsciences (natural science and humanities) produces a significant interaction between year and type of discipline, $F(1, 119) = 5.96$, $p < .02$.
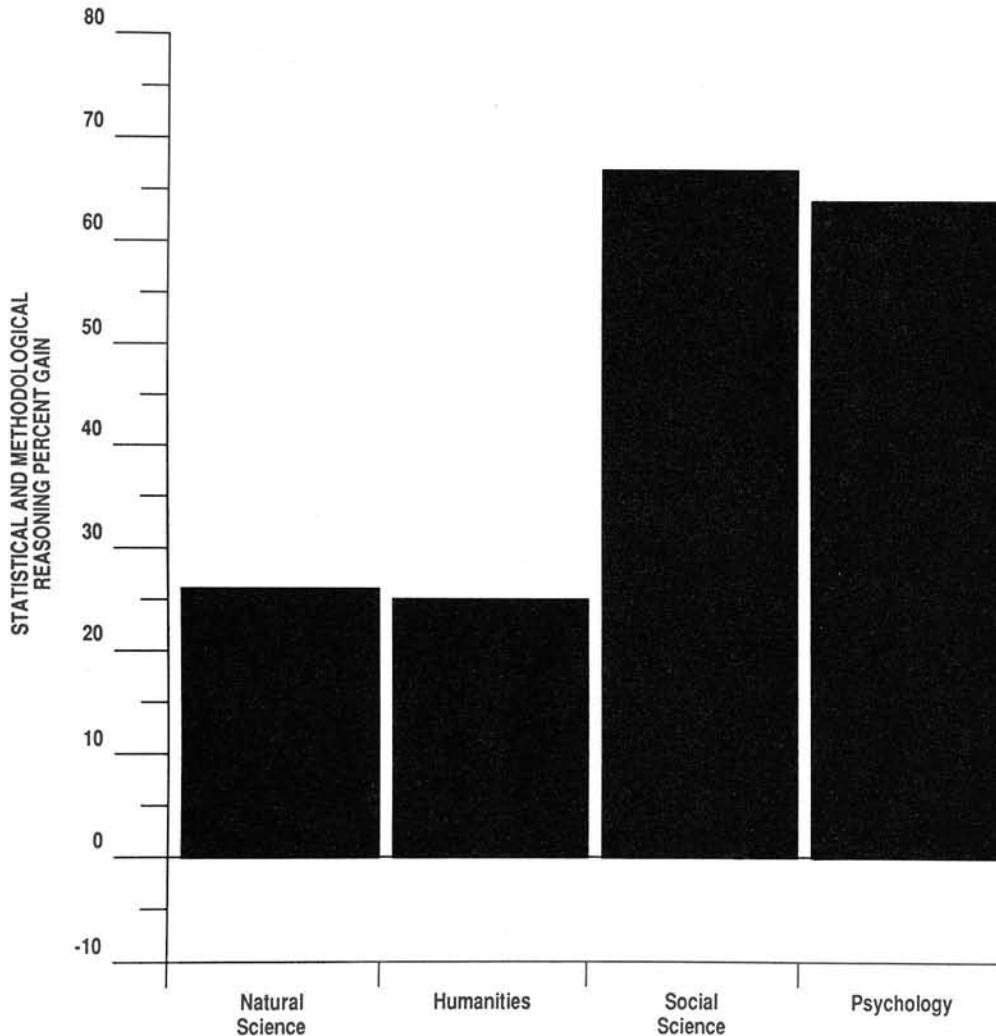
Figure 2. Change scores for statistical and methodological reasoning as a function of year and undergraduate studies.

Conditional reasoning. Figure 3 presents change in conditional reasoning for all subjects. As anticipated, students in the natural sciences improved in their ability to reason about conditional problems, $t(28) = 2.63$, $p < .02$. Surprisingly, humanities students also improved, $t(38) = 3.41$, $p < .003$. In contrast, psychology and social science students did not improve in conditional reasoning, both $ts < 1$. The interaction between year and discipline was significant, $F(3, 117) = 2.64$, $p = .052$.[3]

Correlations between Courses Taken and Improvements in Reasoning

We were able to examine correlations between the number of courses taken of various types and improvement scores (posttest score minus pretest score) on the three types of reasoning skills. Sample sizes are reduced because of missing transcript data. The only significant correlations were between (a) statistical–methodological reasoning improvement and number of statistics courses taken, and (b) conditional reasoning improve-

ment and number of math and number of computer science courses taken.

The correlation between statistical–methodological reasoning improvement and number of statistics courses taken was .17 ($p < .05$) in the overall sample ($N = 103$). The correlations for psychology, social science, and natural science students were .23, .23, and .28, respectively. Only 15% of humanities students

---

[3] The mean percent correct for the pre- and posttests on verbal reasoning was, respectively, 42 and 49 for natural science, 46 and 52 for humanities, 50 and 48 for social science, and 45 and 47 for psychology subjects. The mean percent correct for the pre- and posttests on statistical–methodological reasoning was 26 and 33 for natural science, 27 and 34 for humanities, 23 and 39 for social science, and 25 and 41 for psychology subjects. The mean percent correct for the pre- and posttests on conditional reasoning was 28 and 46 for natural science, 28 and 44 for humanities, 32 and 31 for social science, and 35 and 35 for psychology subjects.
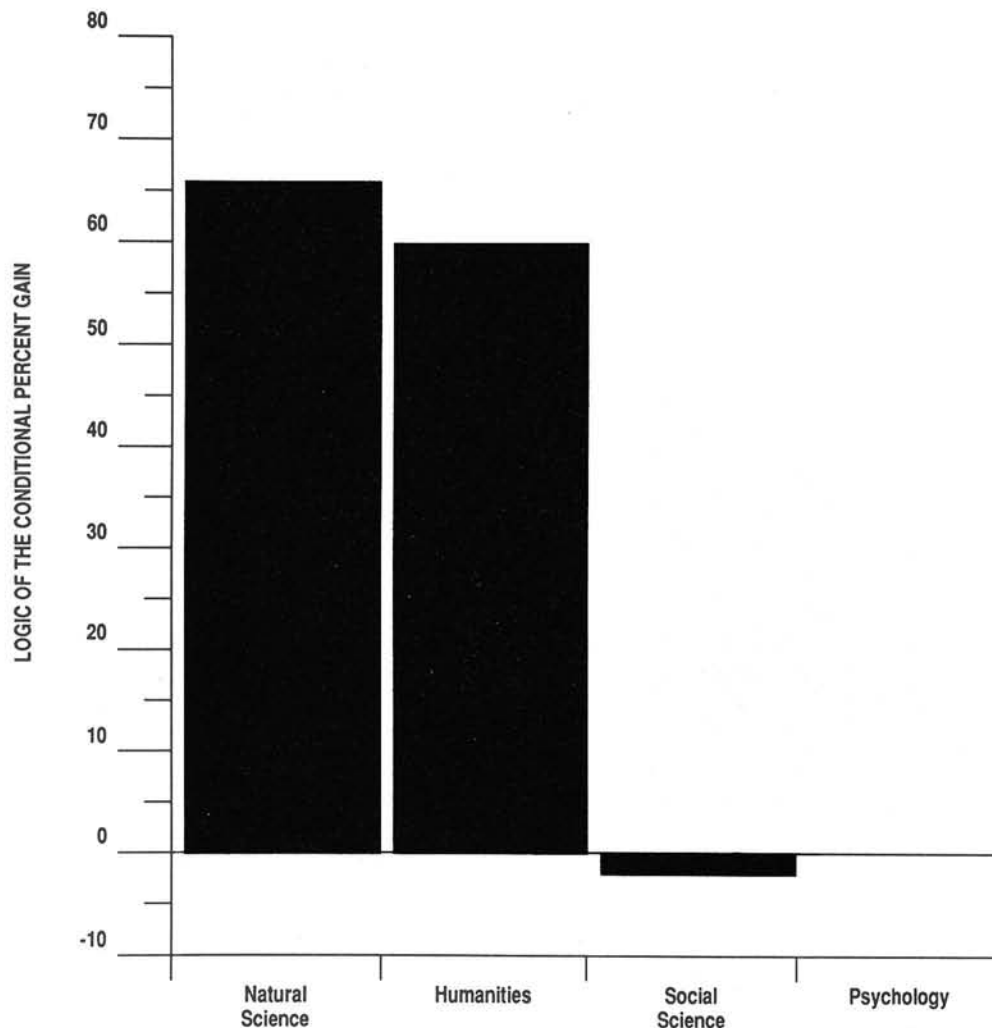
Figure 3. Change scores for conditional reasoning as a function of year and undergraduate studies.

took any statistics courses, and the sample was therefore regarded as too restricted to compute a correlation.

The correlations between conditional reasoning improvement and number of math and number of computer science courses taken were .31 ($p < .002$) and .27 ($p < .005$), respectively, in the overall sample. The correlation between conditional reasoning improvement and math courses taken just for the natural science majors, who took most of the math courses, was .66 ($p < .001$). The correlation between conditional reasoning improvement and computer science courses taken again just for the natural science students (who took the bulk of the computer science courses) was .45 ($p < .02$).

The correlation between conditional reasoning improvement and computer science courses taken was largely spurious because of the correlation between taking math and computer science courses. For the overall sample, the correlation between conditional reasoning improvement and number of math courses taken (partialing out number of computer science courses taken) was .20 ($p < .03$). The same correlation for computer science courses taken (partialing out math courses taken)

was .12 (*ns*). For the natural science group, the correlation between conditional reasoning improvement and number of math courses taken (partialing out number of computer science courses taken) was .55 ($p < .005$). The same correlation for computer science courses taken (partialing out math courses taken) was −.11 (*ns*).

Correlations Between Admission Test Scores, GPA, and Improvements in Reasoning

We were able to examine relations between verbal and quantitative SAT scores and grade point average (GPA), on the one hand, and improvements in the three kinds of reasoning, on the other. The possible patterns of data were many, of course, but we expected a "rich gets richer" pattern. That is, on the basis of a frequent pattern of differential gain found in educational research, we believed it likely that subjects with higher SATs and GPAs would also show the largest improvements in the various reasoning skills. Hence, we expected a significant percentage of strong, positive correlations.

The data did not support this prediction. For the overall sample, only two of the possible nine correlations approached significance. Quantitative SAT score was marginally and positively correlated with improvement in statistical–methodological reasoning, $r(96) = .16$, $p < .06$, and verbal SAT score was marginally and positively correlated with improvement in verbal reasoning, $r(96) = .16$, $p < .06$. Although some correlations for particular majors were higher than these, the pattern was inconsistent and not interpretable.

Discussion

These results are consistent with the view that reasoning can be taught and with the presumption that different undergraduate disciplines teach different kinds of reasoning to different degrees.

Undergraduate training in psychology, and, more generally, the social sciences, can affect statistical and methodological reasoning about a wide range of problems to a very large degree. Students majoring in the natural sciences and the humanities did not improve in statistical–methodological reasoning nearly as much but still showed marginally significant improvements.

The results also indicate that the natural sciences and humanities can affect conditional reasoning. Undergraduate training in the social sciences does not seem to affect this kind of reasoning. We believe that the latter type of training had little effect because, unlike graduate training in the probabilistic sciences, undergraduate training does not emphasize types of causality and the methodologies that are appropriate to each type. The positive effect of the natural sciences in improving conditional reasoning appears to be due to the greater number of math courses natural science students take. We discovered that the more mathematics courses natural science majors took, the more they improved on conditional reasoning, a finding that supports previous work by Jackson and Griggs (1988) indicating that the proof-by-contradiction schema generalizes to problems outside mathematics. We lack an explanation for the gain in conditional reasoning by humanities students.

With the relatively minor gains in verbal reasoning as a backdrop, the percent gains in statistical–methodological reasoning by the social science students, and the percent gains in conditional reasoning by the natural science and humanities students, are quite large. Subjects showed between 60% and 70% improvement, on average, on these classes of problems.

The findings regarding statistical–methodological reasoning seem especially significant. We view these reasoning skills as important for an ability to think critically (see also Ennis, 1987). When deciding what to believe or how to act in everyday life, people must often reflect on the available data and determine their meaning. Accurate assessments frequently require consideration of statistical and methodological questions.

The fact that an undergraduate education can change the way people (a) think about uncertainty in everyday-life events, (b) assess the reliability of information they have before them, and

(c) solve logical problems lends support to the premodern formal discipline hypothesis. Our results show that, contrary to much contemporary theorizing about reasoning, highly general inferential rules can be taught and that the quality of people's everyday reasoning can be improved. Given that universities attempt to hone critical abilities and see this in fact as one of their major tasks, our results paint an optimistic picture.

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., Morris, M., & Nisbett, R. E. (1990). *Schemas for causal deduction.* Unpublished manuscript, University of California, Los Angeles.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: Freeman.

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.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books. (Original work published 1955)

Jackson, S. L., & Griggs, R. A. (1988). Education and the selection task. *Bulletin of the Psychonomic Society, 26,* 327–330.

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.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43,* 431–442.

Newell, A. (1980). One final word. In D. T. Tuma & F. Reif (Eds.), *Problem solving and education: Issues in teaching and research* (pp. 175–189). Hillsdale, NJ: Erlbaum.

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, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

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: Seiler.

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

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247–261, 384–395, 553–564.

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

(Appendix follows on next page)

Appendix

Examples of Items Used

In one of the "scientific-law-of-large-numbers" problems, for example, subjects were told about a study showing that noise harms plants. In the study, two identical coleus plants were transplanted from the same greenhouse and grown under identical conditions, except that the first plant was exposed to about 100 dB of noise (approximately the same as a person would hear while standing on a busy subway platform) while the other plant was grown in quiet conditions. After 1½ weeks of continuous exposure, only the sound-treated plant wilted. A recognition of the law of large numbers was shown by indicating that a larger number of coleus plants should have been used in each of the two conditions. An everyday-life problem to which the regression principle could be applied was identical to one used by Fong, Krantz, and Nisbett (1986): Subjects were asked why rookies of the year in major-league baseball tend not to do as well in their second year. Understanding of the regression principle was demonstrated by indicating that these rookies' performances are unusually good, even for them, and thus it is unlikely that they will have two such unusually good years in a row.

A "scientific-attrition" problem described a study conducted to assess the effects of a weight-loss club. All new members who joined the club during a 1-year period had their weight recorded over time. Out of 138 people who started to attend meetings, 81 kept attending for at least 3 months. In fact, the average amount of weight lost by these people was 14.7 pounds. Appreciation of the attrition principle was indicated by answering that those who enroll in the program and stick to it will on the average be better off, although it is impossible to say whether the techniques used in the program are generally effective. An everyday-life question tapping sample bias effects told of a television debate about a temporary proposal that Alberta, Canada was consider-

ing: the taxing of all people with incomes above $18,000 to deal with a budget imbalance. The debate was followed by a viewer telephone poll in which, for 50 cents, viewers could register their opinion by calling one of two numbers. The final result showed that 64% of the callers wanted the proposal to be defeated, whereas 36% wanted it to pass. Subjects were asked what they thought the results of the election 3 days later would be. Awareness of sample bias was indicated by the answer "I am very uncertain about how the election came out." Note that this question bears a resemblance to the famous poll predicting Landon to beat Roosevelt in the 1936 U.S. presidential election. Because the sample came from lists of owners of automobiles and telephones, the "poor person's candidate" was not given a fair chance.

A "causal schema–conditional" problem read, "As part of your job as a quality control inspector at a shirt factory, you have the task of checking fabric and washing instruction labels to make sure that they are correctly paired. Fabric and washing instruction labels are sewn back to back. Your task is to make sure that all silk labels have the 'dry clean only' label on the other side." Subjects were shown four labels ("machine wash in warm water," "silk," "cotton," and "dry clean only") and were asked to turn over only those labels they would have to check to be sure the labels were correct. The answer is that one must turn over the "machine wash in warm water" label to make sure the other side does not say "silk," and the "silk" label to make sure the other side does not say "machine wash in warm water."