Articles

Teaching Reasoning

Richard E. Nisbett, Geoffrey T. Fong,
Darrin R. Lehman, Patricia W. Cheng

Twentieth-century psychologists have been pessimistic about teaching reasoning, prevailing opinion suggesting that people may possess only domain-specific rules, rather than abstract rules; this would mean that training a rule in one domain would not produce generalization to other domains. Alternatively, it was thought that people might possess abstract rules (such as logical ones) but that these are induced developmentally through self-discovery methods and cannot be trained. Research suggests a much more optimistic view: even brief formal training in inferential rules may enhance their use for reasoning about everyday life events. Previous theorists may have been mistaken about trainability, in part because they misidentified the kind of rules that people use naturally.

Do people use abstract, domain-independent inferential rules to think about everyday events? Can reasoning be improved by formal instruction in the use of inferential rules? Historically, the answer to both questions was yes. That reasoning makes use of inferential rules and that these rules can be taught by formal discipline shaped a position endorsed by most educators through the end of the 19th century.

In the 20th century two different positions have countered the formal discipline view. The first is that there are no inferential rules, but only highly domain-specific empirical rules dealing with concrete types of events. First presented by Thorndike (*1*), this position is still endorsed by many theorists today. For example, Newell (*2*) stated, "The modern . . . position is that learned problem-solving skills are, in general, idiosyncratic to the task." A second view is that people do use abstract inferential rules, but that these cannot be taught to any significant extent. Instead, such rules are induced by every individual in the normal course of development and cannot be improved by instruction. First put forth by Piaget (*3*), this view is highly influential in cognitive and developmental psychology and in education.

We propose an alternative view that is close to the pre-20th-century one: people do make use of inferential rules and these can be readily taught. In fact, rules that are extensions of naturally induced ones can be taught by quite abstract means. This description does not apply to formal, deductive logical rules or to most other purely syntactic rule systems, however. Instead, the types of inferential rules that people use naturally and can be taught most easily are a family of pragmatic inferential rule systems that people induce in the context of solving recurrent everyday problems (*4*). These rule systems are abstract inasmuch as they can be used in a wide variety of content domains, but their use is confined to certain types of problem goals and particular types of relations between events. They include "causal schemas" (*5*), "contractual schemas," such as the rules underlying permission and obligation in the social sphere, and "statistical heuristics," used in the evaluation of evidence, such as qualitative, intuitive versions of the law of large numbers.

We review briefly the history of the formal discipline notion, summarize the views of 20th-century psychologists who opposed it, present evidence from our research that people reason in accordance with abstract inferential rules, present experimental evidence from the laboratory showing that people can be trained to enhance their use of inferential rules for solving everyday problems, and present evidence from studies of higher education showing that something akin to formal discipline is a reality—reasoning can be taught.

Formal Discipline and Its Critics

Plato stated the doctrine of formal discipline, which holds that the study of abstract rule systems trains the mind for reasoning about concrete problems. Plato wrote that the study of arithmetic and geometry was especially effective in improving reasoning, ". . . [E]ven the dull, if they have had arithmetical training, . . . always become much quicker than they would otherwise have been. . . . We must endeavour to persuade those who are to be the principal men of our state to go and learn arithmetic. . ."(*6*).

Later, Roman philosophers added the study of grammar and the training of the faculty of memory to the list of exercises that were useful for formal discipline. The medieval scholastics then added an emphasis on logic, particularly the syllogism. Finally, the humanists added to the previous list the study of Latin and Greek (*7*).

The formal discipline view ultimately became so extreme that a mid-19th-century educator was able to advocate the teaching of a content field strictly for its discipline or exercise properties (*8*):

> My claim for Latin, as an Englishman and a . . . teacher, is simply that it would be impossible to devise for English boys a better teaching instrument. . . . The acquisition of a language is educationally of no importance; what is important is the process of acquiring it. . . . The one great merit of Latin as a teaching instrument is its tremendous difficulty.

One of the first endeavors of the new discipline of psychology was to provide experimental research that cast doubt on the formal discipline concept. The most effective antagonist was Thorndike, who undertook a program of empirical research on transfer of training effects that remains impressive by today's standards. Thorndike rarely found strong transfer effects. He reached the conclusion that the degree of transfer was a function of the number of identical elements in common to the target task and the trained task, where identical elements were defined at the level of relatively concrete

---

R. E. Nisbett is professor of psychology at the University of Michigan, Ann Arbor, MI 48106; G. T. Fong is assistant professor of psychology at Northwestern University, Evanston, IL 60201, and visiting assistant professor of psychology at Princeton University, Princeton, NJ 08544; D. R. Lehman is assistant professor of psychology at the University of British Columbia, Vancouver, BC, Canada V6T 1Y7; and P. W. Cheng is assistant professor of psychology at the University of California, Los Angeles, CA 90024.

features and relations. By taking this position of extreme domain specificity, Thorndike (*9*) was as pessimistic about training effects as previous thinkers had been optimistic.

> Training the mind means the development of thousands of particular independent capacities. . . . The amount of identical elements in different mental functions and the amount of general influence from special training are much less than common opinion supposes.

Thorndike's work was enormously effective in destroying the theoretical rationale for the 19th-century curriculum, which consisted mostly of languages and mathematics and other subjects deemed useful for formal discipline and which largely excluded the natural sciences and other fields because of their emphasis on mere content. But Thorndike's work was really more effective than it should have been, inasmuch as it rarely dealt with reasoning per se. Instead, he studied transfer of training from tasks such as canceling letters in a written message to canceling parts of speech, and estimating areas of rectangles of one size and shape to estimating areas of rectangles of another size and shape. In more recent work in the problem-solving tradition, however, the same conclusion has been reached on the basis of tasks that most people would classify as reasoning tasks. For example, solution of the "Towers of Hanoi" problem (requiring people to move objects from one location to another while preserving size relations among the objects at every location at every point during the move) does not generalize to other, formally identical problems (*10*).

The major psychologist to study reasoning in the period between Thorndike's work and the revolution in thought in the 1960s that saw the shift from a behaviorist viewpoint to an information-processing one was Piaget. Piaget, agreeing with the classical philosophers that everyday reasoning is governed by abstract rules, argued that much reasoning is hypothetico-deductive, handled for the most part by what he called "propositional operations" and by "formal operational schemes." The former are operations in deductive logic; the latter, methods whereby propositional operations are applied to reasoning situations that occur with great regularity in the environment. Formal operations include the probability scheme, an elaborate set of rules for applying probabilistic concepts to uncertain events, such as those produced by randomizing devices.

In Piaget's view, both the propositional operations and the formal operational schemes develop in early adolescence. Before that, children possess only various concrete operations for solving problems, that is, problem-solving rules that are tied to particular content domains. Although Piaget argued that people use inferential rules, he insisted that teaching rules at a high level of abstraction was not possible and that teaching rules in several concrete domains could not accelerate their acquisition in an abstract form. He argued that learning of such rules was dependent on spontaneous cognitive development resulting from active self-discovery.

Piaget's views about both propositional and formal operations have been undermined by recent research (*3*). Piaget's view of propositional operations is cast into doubt because even adults accept invalid arguments when their world knowledge encourages it. For example, given, "All oak trees have acorns. This tree has acorns. Is it an oak?", many adults insist it must be. Defenders of Piaget point out that errors of this type may reflect only vagaries in the interpretation of the arguments, such as the addition or omission of premises due to implicit knowledge. Our knowledge of oaks and acorns, for example, may invite the assumption that only oaks have acorns. Thus, although the conclusion that the tree is an oak is fallacious according to formal logical principles (it is an instance of the fallacy of "affirming the consequent"), it is actually valid within the context of the invited assumptions.

Although it is reasonable to assume that people often do make mistakes because of invited inferences, such interpretive mistakes cannot account for errors produced by college students on deductive reasoning problems in which arbitrary symbols and relations are used. The best known of these problems is Wason's selection task (*11*). In this task subjects are informed that they will be shown cards that have numbers on one side and letters on the other, and are given a rule such as, "If a card has a vowel on one side, then it has an even number on the other." Subjects are then presented with four cards, which might show A, B, 4, and 7, and are asked to turn over only those cards that they have to in order to determine whether the rule holds. The correct answer in this example is to turn over the cards showing A and 7. The logical rule used in such problems is a conditional, "if p then q," and the relevant cases 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 then it must be established that it is also not the case that p). It has been shown with a wide variety of subject populations that when people are presented with such problems with arbitrary relations and no clear semantic interpretation, only a small minority can produce the correct answer.

In the face of evidence that people sometimes seem incapable of using formal logical principles, a view similar to Thorndike's extreme domain specificity has arisen. Since subjects can solve concrete problems in realistic contexts but often cannot solve arbitrary problems, the assertion is sometimes made that all reasoning takes place by domain-specific rules and that more abstract rules play no part in everyday reasoning (*12*).

Piaget's assertion that people use formal operational schemes such as the probability scheme comes from his work showing that even children can use intuitive versions of the laws of probability to predict the behavior of randomizing devices (*13*). But the assumption that these rules generalize to everyday events has been buffeted by recent work, especially that of Kahneman and Tversky and others studying problem-solving heuristics (*14, 15*). These investigators have shown that for many inductive reasoning tasks that require statistical principles, such as the law of large numbers is that the base rate or regression principle, and the conjunction principle, people often fail to use such rules. For example, in one study (*14*) it was shown that subjects did not recognize that a deviant ratio of male to female births was more likely at a hospital with 15 births per day than at a hospital with 45 births per day.

In another study (*14*) subjects were told that 100 people, all of whom were either lawyers or engineers, had been interviewed by psychologists. The subjects were to read what they were told were thumbnail descriptions of the people written by psychologists and asked to guess whether each person described was a lawyer or an engineer. Some subjects were told that the sample of 100 consisted of 70 lawyers and 30 engineers and some were told that the sample consisted of 30 lawyers and 70 engineers. Subjects based their guesses as to occupation almost exclusively on the similarity of the description to their stereotypes for lawyer versus engineer and were little influenced by the frequency of engineers and lawyers in the sample. This was true even when the description was designed to be nondiagnostic with respect to occupation, that is, not suggestive either of an engineer or a lawyer. It should be noted that performance typically is not improved on such problems by various manipulations designed to encourage good work, including monetary incentives for correct answers. Such results have suggested to some that any statistical rules that people may have for the behavior of randomizing devices may not exist at a sufficiently general level to make contact with everyday life problems not involving such devices.

Whether or not these rules exist at a general level, however, both Piaget and theorists who espouse a position of extreme domain specificity are in agreement that teaching abstract rules should be ineffective and that training in a given domain should produce little transfer to other domains. We attempt to show that this pessimistic view of the trainability of inferential rules is mistaken.

Teaching Statistical Reasoning

Our initial work on the use and trainability of inferential rules focused on a set of statistical rules that are derivable from the law of large numbers. We and our colleagues have found that people reason in accordance with the law of large numbers in a wide range of tasks and domains. For example, generalization often proceeds in accordance with the principle that larger samples are required when generalizing about populations that are more variable with respect to a given attribute than when generalizing about populations that are less variable (*16*).

We did, however, find substantial domain specificity of the use of the law of large numbers. There was a hierarchy of usage such that subjects were very likely to use the law of large numbers for reasoning about the behavior of randomizing devices, less likely to use it for reasoning about objectively measurable events such as athletic performance and job and academic achievement, and unlikely to use it for reasoning about subjective events such as judgments about a person's friendliness or honesty (*17*). For example, subjects understood that a small sample of a slot machine's behavior is a poor guide to its behavior in general. They were less likely to apply the law of large numbers to a small sample of an athlete's behavior (for example, to assume that performance at a tryout might not be typical of general ability) and still less likely to apply the law of large numbers to social behavior (for example, to assume that friendliness expressed over a brief time period might not be typical of a person's friendliness in general).

It seems to us that subjects' failure to use the law of large numbers reflects not so much the lack of a general rule, but rather the difficulty of seeing its applicability to events of various kinds. Randomizing devices are by definition those whose behavior cannot be understood by application of causal laws; thus causal laws and other rules do not seriously compete with statistical rules for the "right" to analyze the problem. Objectively measurable events such as sports contests and various achievements normally are understood by the use of various rules about causality, but such events are also sufficiently "codable" (often using literally numerical codes) that people can apply a formal rule such as the law of large numbers. But for purely subjective events, it can be difficult to define the appropriate units for events or to code them on the same scale. Whereas it is possible to directly compare Bill's baseball or sales performance to Sam's with respect to a clear metric (for example, batting average, dollar volume, and so on), it is normally not possible to compare Bill's friendliness or honesty or conscientiousness to Sam's with the use of a clear metric. (What metric would one use to compare Bill's and Sam's friendliness? Smiles per minute?)

One implication of the codability interpretation of domain specificity is that manipulations of the codability of events would be expected to affect people's ability to apply the law of large numbers. And in fact, Nisbett *et al.* (*15*) found that when they made it easier to code events in problems in such a way that their inherent uncertainty was made more apparent, or simply when the events were made easier to compare by suggesting a unit of comparison, a higher proportion of subjects applied the law of large numbers to the problems.

Another finding consistent with the view that people actually possess an abstract version of the law of large numbers is that subjects often justified their statistical answers, across problems having widely varying content, by invoking quite abstract versions of the law of large numbers. This fact demonstrates that people do understand the rule in the abstract and that they know how to describe correctly how it can be mapped onto a solution for a given problem.

One way of providing evidence that people can use purely abstract rule systems would be to teach the rule system in order to see if people can apply it to a wide range of events for which it is applicable. Two different means of teaching would be persuasive on this point: (i) brief formal instruction in the abstract rule system and (ii) brief instruction in use of the rule system in a single domain. If purely formal instruction is effective across a wide variety of content domains, this would suggest a preexisting rule system and an ability to apply improvements to it to many domains. If training in a given domain generalizes substantially to other domains, this would suggest a substantial ability to abstract the rule, or rather, improvements to it, from particular content and would further imply a preexisting ability to apply uninstructed versions of the rule to a broad range of content.

We have examined the effects of teaching the law of large numbers both in the abstract and with examples drawn from a given broad domain (*18*). Abstract instruction in the law of large numbers consisted of defining the notions of sample, population, and parameter, and illustrating that as sample size increases, the sample usually resembles the population more closely. Training on examples consisted of showing how to use the law of large numbers to solve a number of problems. In one problem, subjects read about the director of a ballet company who uses an audition to select new dancers. "Usually we're extremely excited about the potential of two or three of these young people—a young woman who does a brilliant series of turns or a young man who does several leaps that make you hold your breath. Unfortunately, most of these young people turn out to be only somewhat better than the rest. I believe many of these extraordinarily talented young people are frightened of success." Subjects were invited to appreciate the relevance of the law of large numbers to the problem in addition to the director's exclusively causal view: they were encouraged to think of each ballet dancer as possessing a population of ballet movements and to think of the audition as providing samples of that population; then since the sample is relatively small, the population value might well be different, especially when the sample value is quite extreme. (The measure in this case was simply the director's evaluation of each individual movement.) Examples were drawn exclusively from one of the following broadly defined domains—(i) manifestly probabilistic problems such as those involving random generating devices (for example, a problem in which the first few cards drawn from a shuffled deck in a board game are all from a particular category and subjects must realize that such a small sample result may not be indicative of the overall proportion in the category), (ii) objectively measurable events such as those involving achievements of some kind (for example, a problem about whether the performance of graduates of a particular law school in general could be safely estimated on the basis of the performance of just two particular graduates), or (iii) subjective judgments such as those assessing someone's sense of humor or kindness (for example, a problem in which the subjects must realize that a first impression of a person might not be a good indication of that individual's personality, at least not in comparison to impressions reached after a longer acquaintance).

Some subjects received no training at all, some received abstract rule training only, some examples training only, and some received both rule and examples training (*19*). Each of the two separate training procedures took less than 30 minutes. All subjects were then tested on their ability to apply the law of large numbers to a variety of problems in all three domains. Subjects in one variant of the experiment were University of Michigan students; subjects in another variant were New Jersey homemakers, most with secondary education only, who were paid for their participation. Results from the two variants were entirely comparable and are reported together.

Subjects' open-ended answers to the test problems were coded and two dependent variables were created. "Frequency" was defined as the overall probability that a subject responded to a test problem by incorporating statistical concepts such as sample size, variability, or uncertainty, without regard to whether the subject used these concepts properly. "Quality" was defined as the conditional probability that a subject gave a "good statistical answer," that is, one in which the subject had made appropriate use of statistical principles, identifying the metric, the sample, the population, and their correct relation to one another.

Three results support the view that people possess an abstract version of the law of large numbers and that improvements to it can transfer to a wide range of problem content. First, purely abstract rule training produced improvement in both the frequency and the quality of statistical answers. Second, the abstract rule training effect was substantial across all three problem domains: training improved statistical reasoning for problems that people rarely think of in terms of probability just as much as it did for problems that people almost always think of in probabilistic terms. On average, rule training increased the percentage of statistical answers from 42 to 56 and increased the percentage of statistical answers judged by coders to be of high quality from 54 to 67 (*20*). Third, training on examples readily generalized to domains very different from the trained domain. Indeed, generalization across domains was literally as great as generalization within domains. This seems hard to explain without assuming that instruction in the particular examples served to improve the abstract rule system underlying solution of all problems.

A follow-up study with Northwestern University students examined the effect of "examples" training with narrower domains (*21*). Subjects were taught how to apply the law of large numbers either to a variety of sports problems or to a variety of problems having to do with ability testing. For example, in the sports training condition, subjects were asked to explain the fact that, after the first 2 weeks of the major league baseball season, the top batter typically has an average as high as .450, yet no batter has ever had as high an average as that at the end of the season. After subjects had tried their hands at the problem (usually providing exclusively causal, though not necessarily wrong, answers such as "the pitchers make the necessary adjustments"), they were shown the investigators' analysis which pointed out that, whatever causal factors might be involved, it is also relevant to note that 2 weeks provides a relatively small sample of a batter's ability and that batting averages that are highly discrepant from the average should therefore be more common than they are with a large sample.

Subjects' performance was then examined either in the trained domain or in the untrained domain, either immediately or after a delay of 2 weeks. Figure 1 presents the statistical reasoning score, that is, the average quality of the answer for each of five different problems. The remarkable domain independence of training effects is evident for subjects tested immediately. There was no significant advantage to being trained in a given domain: performance was as good for the untrained domain as for the trained domain. After a delay, there was a substantial degree of domain specificity. Even then, however, there remained a significantly greater ability to apply the rule across domains after the delay than without any training at all.

It should be noted that the full degree of improvement initially observed in the trained domain is still observed after 2 weeks. It is unlikely either that the full retention of training within a domain or the improved performance across domains is due to retention of problem details and consequent mapping of details from the example problems to the new test problems by construction of analogies. In separate studies (*21*), it was found that ability to spontaneously recall details, or even the gist of the training examples, was extremely low after 2 weeks. We suspect that subjects learned improvements to the abstract rule system initially, as well as some specific abilities to apply the rule system to a particular content domain. After a delay, the increments to both general and specific rules were fully retained, and thus performance on the trained domain was as high as it was initially. Only the increments to general rules could be passed along directly to the untrained domain, and thus performance in the untrained domain was poorer after a delay—though nevertheless better than in the absence of training.

Teaching Logical Reasoning

Since highly abstract statistical rules can be taught in such a way that they can be applied to a great range of everyday life events, is the same true of the even more abstract rules of deductive logic? We can report no evidence indicating that this is true, and we can provide some evidence indicating that it is not.

A purely syntactic rule system provides no information about which inferences among infinitely many valid inferences are useful. For example, given the statement, "If a burglar is breaking into one's house, then one should call the police," it is valid though hardly useful to infer, "If it's not the case that one should call the police, then a burglar is not breaking into one's house." In our view, when people reason in accordance with the rules of formal logic, they normally do so by using pragmatic reasoning schemas that happen to map onto the solutions provided by logic.

One type of pragmatic reasoning schema that could mimic the logic of the conditional is the "causal schema" identified by Kelley (*5*). He proposed that, when making causal attributions, people sometimes rely on very abstract rules about the kind of relations that obtain when the particular type of causality in question is of the necessary sort, the sufficient sort, or neither. We have found support for Kelley's contention. For example, people use different procedures for establishing whether a causal hypothesis is true depending on whether they assume that the type of causality they are examining is necessary and sufficient, necessary but not sufficient, sufficient but not necessary, or neither necessary nor sufficient (*22*). The checking procedure for one of the causal schemas is identical to that for the conditional, namely, the sufficient but not necessary type of cause. To falsify such a hypothesis, one may either establish that the effect q is not present when the putative cause p is present, or establish that the cause p is present when the effect is not. The checking procedures for one of the other causal schemas, namely, the necessary and sufficient schema, are the same as those for the biconditional.

Another type of pragmatic reasoning schema is what might be called the contractual schema. The concept of a permission establishes that one will not be allowed to do action p unless permission q has been obtained. The concept of an obligation establishes that if situation p occurs then one is obliged to do action q. As it happens, the procedures for checking whether an infraction of a permission or obligation has occurred are the same as those for checking whether the conditional obtains. For example, one must establish that q has occurred (permission has been obtained) when one finds that action p has been carried out, and one must establish that action p has not been carried out when one knows that q (permission) has not occurred.

If people actually make use of schemas such as those for permis-

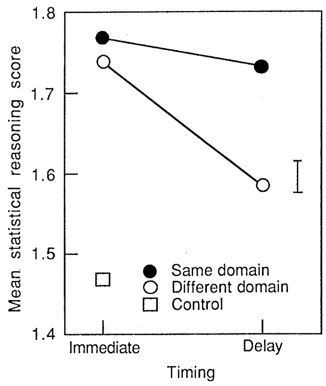

Fig. 1. Mean statistical reasoning score as a function of training and problem domain. Vertical bar represents one standard error of the mean (*N* = 231).

sion and obligation, then they should be able to solve problems whose checking procedures are identical with the logic of the conditional when they are encouraged to apply such schemas. Several studies indicate that this is true (*23*). For example, many subjects who were given no context were unable to solve a simple conditional problem about deplaning airport passengers who were required to show a form. They were required to check whether the rule, "If the form says 'entering' on one side, then the other side includes cholera among the list of diseases," was violated by different instances. Almost all subjects readily solved the problem, however, when they were provided with a "permission" rationale, to wit, they were told that the form listed diseases for which the passenger had been inoculated, and a cholera inoculation was required to protect the entering passengers from the disease.

Another study examined the effects of training in the logic of the conditional on subjects' ability to solve both arbitrary conditional problems and problems evoking permission and obligation schemas (*24*). Neither abstract rule training nor examples training showing how to use the conditional for solving particular concrete problems was effective. It is not the case that the training procedures were inherently useless, however, because when subjects were given both of the training procedures, this resulted in a very significant improvement in their performance. This is in sharp contrast to the statistical training studies, where both abstract rule training and examples training were effective alone. We believe that abstract logical training by itself was ineffective because the subjects had no preexisting logical rules corresponding to the conditional. (Or, more cautiously, any such rules are relatively weak and not likely to be applied in meaningful contexts.) Showing subjects how to use the rule to solve example problems was ineffective for the same reason. It was only when subjects were given both types of training that they could make use of either. The training procedures and problems used were far from exhaustive of reasonable approaches to teaching the conditional, but the results suggest that it may be difficult to teach logical rules by the straightforward techniques used to teach what we call pragmatic inferential rules.

In marked contrast to the effects of teaching abstract logical rules, training in the obligation schema was highly effective in improving performance both on problems that were suggestive of the obligation rule and on arbitrary problems that could be mapped onto it. We believe that this is the case because the obligation-based training could be attached to a preexisting knowledge structure whereas the purely syntactic training could not.

Our work suggests that the formal discipline view may well be correct in essence, but that it has misidentified the knowledge structures that underlie reasoning about everyday life events. There are abstract rules, and these can be trained abstractly. The rules may not be those of formal logic, however, but instead may be pragmatic inferential rules having to do with particular types of relationships and inferential goals. These structures are more specific than logical rules, but they are abstract in that they are not bound to any content domain.

Implications for Higher Education

It appears that inferential rules can be taught in the laboratory, and taught in such a way that they are reasonably enduring. Does higher education serve to teach these inferential rules in the same way, with similar or even greater durability? The answer appears to be yes, for both statistical rules and pragmatic reasoning schemas such as causal schemas and contractual schemas.

Teaching statistics. Fong *et al.* (*18*) examined the effects of differing amounts of statistical education on answers to a problem asking subjects to explain why a traveling saleswoman is typically disappointed on repeat visits to a restaurant where she experienced a truly outstanding meal on her first visit. Subjects who had no background in statistics almost always answered this problem with exclusively nonstatistical, causal answers such as "maybe the chefs change a lot" or "her expectations were so high that the food couldn't live up to them." Subjects who had taken one statistics course gave answers that included statistical considerations, such as "very few restaurants have only excellent meals, odds are she was just lucky the first time," about 20 percent of the time. Beginning graduate students in psychology, who had taken one to three courses in statistics, gave statistical answers about 40 percent of the time. Doctoral-level scientists at a research institution gave statistical answers about 80 percent of the time. (Though we do not wish to create the impression that these scientists would necessarily think in statistical terms so often in real life contexts! In this case, performance in the laboratory undoubtedly outstrips competence in the world.) Training affected the quality of statistical answers as much as it did their frequency. Subjects with but one statistics course rarely gave an answer that did much more than just point to the chance nature of the quality of any one meal; subjects with many courses often defined meals as the sampling unit, defined the population as all possible meals in the restaurant, invoked the notion of variability in meal quality within a restaurant, and so on.

The study just described was correlational. Another study of higher education by Fong *et al.* (*18*) was experimental, and replicated the effects. They conducted a telephone survey of opinions about sports. The subjects were males who were enrolled in an introductory statistics course at the University of Michigan. Some subjects were randomly selected and "surveyed" during the first 2 weeks of the term, the others at or near the end of the term. In addition to filler questions on National Collegiate Athletic Association rules and National Basketball Association salaries, subjects were asked questions for which a statistical approach was relevant. For example, subjects were asked to explain why the rookie of the year in baseball usually does not perform as well in his second year. Most subjects answered this question in a purely nonstatistical way, invoking causal notions such as "too much press attention" and "slacking off." Some subjects answered the question in a partly or completely statistical way, for example, "there are bound to be some rookies who have an exceptional season; it may not be due to any great talent advantage that one guy has over some of the others—he just got a particularly good year." The statistics course markedly increased the frequency and quality of statistical answers to this question and to two of four other questions that were asked.

Teaching logic. If it is correct that systems of formal deductive logic are alien and hard to teach, then one might expect that even an entire course in formal logic would have little effect on students' ability to deal with problems that can be solved by the use of the conditional or biconditional. To test this, Cheng *et al.* (*24*) examined the effects

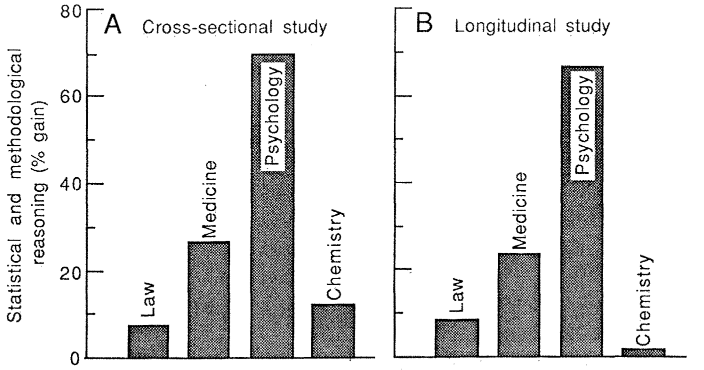

Fig. 2. Percentage of change in statistical and methodological reasoning score after 2 years of study as a function of graduate discipline. (**A**) The cross-sectional study examined first-year students and simultaneously enrolled third-year students. Sample sizes for first-year students were law, 213; medicine, 127; psychology, 25; and chemistry, 31. The sample sizes for third-year students were 50, 48, 33, and 26, respectively. (**B**) The longitudinal study examined the same students at the beginning of their first year and at the beginning of their third year. Sample sizes were law, 77; medicine, 87; psychology, 24; and chemistry, 18.

of introductory logic courses given at two different universities, one a highly selective state university and one a less prestigious branch of the same university. The course at the former university was exclusively concerned with teaching formal deductive systems. The course at the latter also dealt with informal fallacies. Quite different texts were used in the two courses. But both courses covered topics in propositional logic, and indeed, both built from an initial foundation on the logic of the conditional, including the biconditional.

A pretest consisting of both meaningful selection problems (for example, ones inviting a causal or permission interpretation) and arbitrary ones (for example, the original Wason card problem) was given in the first week of class before any discussion of the conditional had taken place. A posttest was given in the final week of the semester. Problems on the pretest and posttest, although embodying the same principles as those taught in both logic courses, did not correspond directly to any problems actually presented in either course. The results provided no statistically significant evidence that formal instruction in logic can improve reasoning performance as measured by the selection task. The mean improvement was only 3 percent. The percentage of biconditional problems solved correctly actually decreased trivially.

Effects of graduate training. Graduate programs provide an excellent opportunity to examine the effects of intensive training in particular types of inferential rules. Different fields emphasize different rule systems, and, unlike undergraduate school, where students are exposed to many different disciplines, the narrow focus of graduate programs might make it possible to show distinct patterns of inferential gains.

Lehman, Lempert, and Nisbett (*25*) studied the effects of 2 years of graduate education in four different fields in which inferential rules are taught that are extensions of the naturally induced pragmatic rules we have identified. Graduate students in these fields were tested on several different kinds of inferential skills: (i) statistical reasoning about both scientific content (such as statistically flawed studies in the natural and social sciences) and everyday life content (such as the sports or restaurant meal problems described above), (ii) methodological reasoning dealing with different types of confounded variable problems, for example, self-selection problems (*26*), sample bias problems, and inferential uses of control groups (*27*), for both scientific content and everyday life content, and (iii) ability to solve both arbitrary and meaningful problems involving the conditional and biconditional. Four fields at the University of Michigan were examined—psychology, medicine, law, and chemistry. Two different studies were conducted, one with a cross-sectional design (that is, first-year students in each field were compared with third-year students) and one with a longitudinal design (that is, first-year students in each field were tested and, after 2 years of training, tested again). The expectations were that training in the probabilistic sciences of psychology and medicine would result in an enhanced ability to apply statistical and methodological rules to both scientific content and everyday life problems. In addition, because psychology and medicine must deal with all kinds of causal patterns involving necessity and sufficiency, it was expected that training in these fields would also improve ability to solve conditional and biconditional problems. The field of chemistry, dealing as it does primarily with necessary and deterministic causes, was expected to produce little improvement in ability to apply rules for dealing with uncertainty or with the conditional. It was also expected that training in law would produce little improvement in ability to apply rules for dealing with uncertainty. Training in law, however, was expected to produce improvements in the ability to reason about problems that could be solved either by use of the conditional or by use of the biconditional. This is because law students are taught about contractual relations.

Initial differences among the three groups were very slight for all types of reasoning studied. Figure 2 shows that the effects of 2 years of training conformed closely to the pattern just described. The effects of training on ability to use statistical rules and confounded variable rules were marked, for both psychology and medical students, both for scientific problems and for everyday life problems. (The effects for both types of rules were almost identical and results were combined in Fig. 2.) The change for psychology students was particularly great, resulting in approximately an 80 percent increase in ability to apply both types of rules for both studies. The change for medical students was also statistically significant for both studies. Neither law students nor chemistry students improved in reasoning using either statistical rules or confounded variable rules, for either type of content, when studied by either type of design.

For problems involving the logic of the conditional, it may be seen in Fig. 3 that psychology, medicine, and the law were all effective, and about equally so, producing about a 30 percent improvement. Changes for the cross-sectional study were not statistically significant, whereas all changes for the more sensitive longitudinal study were. Chemistry training had no effect on conditional problems in either study.

To increase generalizability, the cross-sectional version of the study was replicated at the University of California at Los Angeles for psychology and chemistry students. The results were similar. Chemistry students showed no gain for statistical or methodological problems or for conditional problems. Psychology students improved in all three.

The Future of Formal Discipline

Taken together, the results of our studies suggest that the effects of higher education on the rules underlying reasoning may be very marked. In fact, the effects may be marked enough to justify the teaching of some rule systems invoking precisely the principles of formal training and general transfer that have long been invoked for logic, grammar, and other formal systems.

Our results also suggest that contrary to the pre-20th-century view of formal discipline, higher education does not train the mind as physical exercise trains the muscles. For example, although law

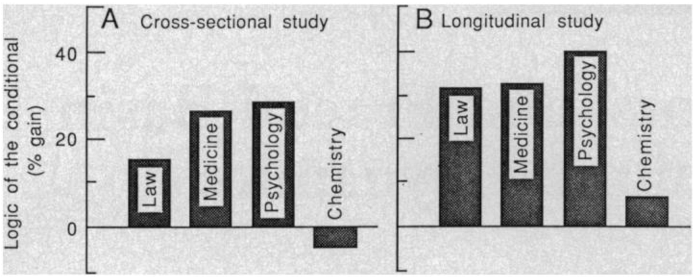

Fig. 3. Percent change in ability to solve problems in the logic of the conditional after 2 years of study as a function of graduate discipline.

students improved on conditional problems, possibly because of training about contractual relations, the improvement did not generalize to statistical rules or confounded variable rules. Thorndike (*1, 9*) was partially correct, after all, in that transfer applies only insofar as there are common identical elements. But the identity lies at a much higher level of abstraction than he suggested, at the level of pragmatic inferential rules such as contractual schemas, causal relations, and the law of large numbers. Furthermore, also contrary to his thesis, transfer does not necessarily occur when problems share identical elements. There is transfer neither for identity as specific as different types of isomorphic Tower of Hanoi problems, nor for identity as general as modus tollens (the contrapositive rule that a logician can use to solve the Wason selection problem). We suggest that transfer in the domain of reasoning may occur only when the identical elements are pragmatic inferential rules.

The results also indicate that Piaget was mistaken in two respects: (i) people may not possess rules of propositional logic at the purely syntactic level (at least in a form such that they are used for meaningful problems), and (ii) it is indeed possible to improve inferential rules through training.

Thus we propose there is such a thing as formal discipline—teaching people how to reason. It should be noted that our optimism is consistent with recent findings indicating that it is possible to train such foundations of reasoning as how to use dimensions to analyze and organize similarities and differences and how to identify the structure of simple propositions (*28*). Our view is also consistent with process-oriented theories of intelligence that emphasize the pragmatic experiential context in which intelligence evolves in the context of everyday problem solving (*29*).

Now that there are some clues about the nature of the rules that people actually use and can be taught, we may be able to proceed more efficiently to identify the ones that are most useful and how they can best be taught.

REFERENCES AND NOTES

1. See, for example, E. Thorndike, *The Psychology of Learning* (Mason-Henry, New York, 1913).
2. A. Newell, in *Problem Solving and Education*, D. Tuma and F. Reif, Eds. (Erlbaum, Hillsdale, NJ, 1980).
3. For a review, see C. Brainerd, *Piaget's Theory of Intelligence* (Prentice-Hall, Englewood Cliffs, NJ, 1978).
→ P. Cheng and K. Holyoak, *Cognit. Psychol.* **17**, 391 (1985); J. Holland, K. Holyoak, R. Nisbett, P. Thagard, *Induction* (MIT Press, Cambridge, MA, 1986).
→ H. Kelley, *Am. Psychol.* **28**, 107 (1973).
6. B. Jowett, Ed., *The Dialogues of Plato* (Oxford Univ. Press, Oxford, 1875), p. 785.
7. L. Mann, *On the Trail of Process* (Grune & Stratton, New York, 1979).
8. B. Tarver, quoted in Mann (*7*), p. 132.
9. E. Thorndike, *Principles of Teaching* (Seiler, New York, 1906), p. 246.
10. J. R. Hayes and H. A. Simon, *Cognitive Theory*, N. J. Castellan, Ed. (Erlbaum, Hillsdale, NJ, 1977).
11. P. Wason, in *New Horizons in Psychology*, B. Foss, Ed. (Penguin, Harmondsworth, England, 1966).
12. R. Griggs and J. Cox, *Br. J. Psychol.* **73**, 407 (1982).
13. J. Piaget and B. Inhelder, *The Origin of the Idea of Chance in Children* (Norton, New York, 1975).
→ A. Tversky and D. Kahneman, *Science* **185**, 1124 (1974).
15. R. Nisbett and L. Ross, *Human Inference* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
16. R. Nisbett, D. Krantz, C. Jepson, Z. Kunda, *Psychol. Rev.* **90**, 339 (1983).
17. C. Jepson, D. Krantz, R. Nisbett, *Behav. Brain Sci.* **6**, 494 (1983).
18. G. Fong, D. Krantz, R. Nisbett, *Cognit. Psychol.* **18**, 253 (1986).
19. An additional control group received brief "placebic" instruction: the law of large numbers was lauded and defined as the superiority of larger quantities of evidence over smaller quantities of evidence, and subjects were exhorted to use the law in solving test problems. This "demand" control group was not different from completely untrained control subjects in either the frequency or quality of statistical explanations produced.
20. Sample size for the rule training group was 69 and for the control group, 68.
21. G. Fong and R. Nisbett, unpublished data.
22. P. Cheng and R. Nisbett, unpublished data.
→ P. Cheng and K. Holyoak, *Cognit. Psychol.* **17**, 391 (1985).
24. ______, L. Oliver, R. Nisbett, *ibid.* **18**, 293 (1986).
25. D. Lehman, R. Lempert, R. Nisbett, unpublished data.
26. An example of a problem involving self-selection was drawn from an editorial in the *New York Times* (6 March 1982, p. 23) in which the Learning of Latin and Greek was urged on the grounds that high school students who had done so were found to score a standard deviation higher on the SAT verbal tests. Subjects able to apply the self-selection rule recognized that the difference would be due largely to initial differences in ability between the sorts of students who would take Latin and Greek and those who would not.
27. An example of a problem in which the ability to apply control group concepts to everyday life is examined is one noting that the Mayor of Indianapolis is under pressure to fire his police chief because crime has increased since the chief began his tenure in office. Subjects able to apply control group concepts recognize that a decision to fire should be based in part on the crime rates in comparable cities over comparable periods of time.
28. R. Herrnstein, R. Nickerson, M. de Sanchez, J. Swets, *Am. Psychol.* **41**, 1279 (1986).
→ E. Hunt, *Science* **219**, 141 (19 → R. J. Sternberg, *ibid.* **230**, 1111 (1985).

30. Supported by NSF grants SES85-07342, BNS84-09198, and BNS-8709892, NIMH grant 1 RO1 MH38466, and grant 85-K-0563 from the Office of Naval Research. We thank K. J. Holyoak, L. Novick, and E. E. Smith for comments on an earlier draft.

". . . AND IT'S LOADED WITH ADDITIVES BANNED BY THE FDA"