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).
14. 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 G. Fong and R. Nisbett, unpublished data.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 P. Cheng and R. Nisbett, unpublished data.
23. 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 D. Lehman, R. Lempert, R. Nisbett, unpublished data.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).
29. E. Hunt, *Science 219*, 141 (1983); 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.

13 The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking About Everyday-Life Events

Darrin R. Lehman
University of British Columbia

Richard E. Nisbett
University of Michigan

Richard O. Lempert
University of Michigan Law School

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

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

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

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

THE 20TH CENTURY CRITIQUE OF FORMAL DISCIPLINE

Although there were objections to the standard curriculum as early as the Enlightenment on the grounds that the rules of mathematics and Latin bear little actual resemblance to the rules necessary to think about most everyday-life events, it was not until the late 19th century that the view came under concerted attack. The attack came from psychologists, and it was utterly effective. In fact, it was one of the first policy victories of the new field. William James was withering in his critique of "faculty psychology," that is, of the view that mental abilities consisted of faculties such as memory, reasoning, and will that could be improved by mere exercise in the way that muscles could. Thorndike (1906, 1913) undertook a program of research, still impressive by modern standards, that showed little transfer of training across tasks, for example, from canceling letters to canceling parts of speech or from estimating areas of rectangles of one size and shape to estimating areas of rectangles of another size and shape. Thorndike declared that "training the mind means the development of thousands of particular independent capacities" (Thorndike, 1906, p. 246). Hence, training in Latin could not be expected to improve people's capacities to perform other very different mental tasks.

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

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

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

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

PRAGMATIC INFERENTIAL RULES

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

Contractual Schemas

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

Cheng and Holyoak (1985) showed that the schema for permissions is useful in performing selection tasks having the form of the Wason card problem. In one of Cheng and Holyoak's experiments, subjects were presented with a selection problem based on an abstract description of a permission situation: "If one is to take action 'A,' then one must first satisfy precondition 'P.' " About 60% of the subjects solved this problem as compared with 20% who solved the original Wason selection task. Cheng and Holyoak also found that providing an explicit purpose for a rule that would otherwise seem arbitrary could serve to cue the permission schema and hence facilitate performance. These findings are incompatible with the extreme domain specificity view stemming from Thorndike's position. The findings also indicate that the selection task, with its arbitrary elements and relations, is difficult because people normally reason using schemas of the permission type rather than the rules of formal logic. When reasoning schemas are invoked, people can solve problems that are formally identical to the card selection task because the schematic rule structures are identical.

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

Causal Schemas

Cheng, Nisbett, and Oliver (1987) have argued that another type of pragmatic reasoning schema, namely causal schemas of the kind defined by Kelley (1971, 1973), may also help people to solve problems that are syntactically identical to the selection problem. Kelley argued that people understand the ideas of necessariness and sufficiency in causality and possess different schemas for checking evidence supporting causal hypotheses that are necessary and sufficient, necessary but not sufficient, and so on. Cheng et al. (1987) found support for this view. They found that subjects tacitly assumed a particular type of causality and then applied evidence-checking procedures appropriate to the type. Interestingly, one of the schemas, namely that for sufficient but not necessary causes, has checking procedures identical to that for the conditional selection task. A hypothesis of the type "A among other things always causes effect B" can be disproved by examining A and finding that not-B was the case, or examining not-B and finding that A was the case. Cheng et al. (1986) found that subjects performed better on selection tasks that had a causal interpretation that might encourage subjects to use the same checking procedures as required by the conditional. The checking procedures for the necessary and sufficient type of causality, as it happens, are the same as for the biconditional *(p if and only if q)* in formal logic. The biconditional requires examining all four cases—*p, not-p, q,* and *not-q.* The other two causal schemas—namely, necessary but not sufficient, and neither necessary nor sufficient—also have distinct checking procedures associated with them. The social sciences, it should be noted, have developed elaborate methodologies for dealing with the completely probabilistic type of causality characterized by the neither necessary nor sufficient type.

Statistical Rules

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

GRADUATE SCHOOL AND FORMAL DISCIPLINE

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

The work also has some clear implications for understanding the effects of education on the way people reason. Because graduate training in particular is highly specialized with respect to the inferential rules emphasized, the possibility exists that different kinds of graduate training produce different effects on reasoning about various everyday-life events. In particular, we would expect that scientific disciplines teach different rules than non-scientific disciplines and that there might be differences within the sciences themselves having to do with whether the field is at base probabilistic or deterministic. We studied two probabilistic sciences, namely psychology and medicine, a nonprobabilistic or deterministic science, namely chemistry, and a nonscience, namely law.

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

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

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

We investigated the effects of different types of graduate education on