Chapter 1

Logic and the Study of Human Reasoning

1.1 Why were psychologists interested in deductive logic?

> The world, in short, was providing not _sensation_ but fodder for our hypotheses.
>
> -- Jerome Bruner*

For many, the degree to which human learning mechanisms can be counted on to produce valid knowledge is the measure of man's rationality. But what characteristics must a learning process have to ensure that the knowledge acquired is valid? Because the generalizations of science afford the closest approach to what we intuitively think of as valid knowledge, psychologists have watched the philosophy of science closely to learn which characteristics of the scientific learning process are epistemologically criterial. If everyday learning can be shown to share these criterial characteristics with its more refined sister, then human rationality is spared.

Ever since Hume, induction has carried a heavy load in psychology while taking a sound epistemological beating. In psychology, it has been the learning theory of choice since the British Empiricists argued that the experience of spatially and temporally contiguous events is what allows us to jump from the particular to the general, from sensations to objects, from objects to concepts. Pavlovian reflexology, Watsonian and Skinnerian behaviorism, even the sensory-motor parts of Piagetian structuralism have been mere essays on the inductive psychology

---

* Bruner, 1984, p. 95, emphasis his.

of the British Empiricists. Yet when Hume, a proponent of inductive inference as a psychological learning theory, donned his philosopher's hat, he showed that induction could never justify a universal statement. Thus, Hume showed that the process by which people were presumed to learn about the world could not ensure that the generalizations it produced would be valid.

Only recently, with the publication in 1959 of Karl Popper's The Logic of Scientific Discovery, has the philosophical beating of psychology's favorite learning theory subsided. Popper argued that although a universal statement of science can never be proved true, it deductively implies particular assertions about the world -- hypotheses -- and particular assertions can be proved false. No number of observed white swans can prove that "All swans are white" is true, but just one black swan can prove it false. Generalizations cannot be confirmed, but they can be falsified, so inductions tested via deductions are on firmer epistemological ground than knowledge produced through induction alone.

This view had consequences for psychologists interested in everyday learning. If one assumes that the evolutionary purpose of human learning is to produce valid generalizations about the world, then surely everyday learning must be some form of Popperian hypothesis testing. On this view, induction still plays an important role in the creation of knowledge -- it is a source of testable hypotheses. But the burden of validating knowledge now falls on deductive logic. Inductive processes might suggest "If P then Q" -- and "If R then Q", "If S then Q", and so on -- but it is our use of deductive logic that causes us to reject the hypothesis if Q turns out to be false when P (or R or S) is true. The repeated application of deductive logic in testing the many inductively produced hypotheses explaining Q lets us hone down the possibilities and zero in on the truth. On this view, human learning processes will produce valid knowledge to the extent that they use deductive logic to falsify hypotheses.

This shifts theoretical priorities in psychology from the study of inductive processes to the study of deductive logic. Either everyday human learning mechanisms usually produce invalid knowledge, or they include algorithms that frequently and spontaneously apply deductive logic in testing hypotheses.

A legion of cognitive psychologists, from Piaget (e.g., Inhelder & Piaget, 1958, p. 254-255) to Bruner (Bruner, Goodnow & Austin, 1956) to Wason & Johnson-Laird (1972) to Fodor (1975) have adopted this Popperian view of hypothesis testing as their model of human learning. To paraphrase Fodor, who was speaking for the field, hypothesis testing is the only theory we've got (Fodor, 1975, Ch. 1).

Deductive logic has another property that was tempting to cognitive psychologists: it is content-independent. The rules of inference of the propositional calculus* generate only true conclusions from true premises, regardless of what the propositional content of those premises is. The propositional calculus is the perfect inference engine for a domain general information processing system: no matter what hypotheses inductive processes feed it, it will output only valid conclusions. The idea that the human mind has algorithms that

---

* the philospher's name for formal propositional logic (Quine, 1950).

instantiate the rules of inference of the propositional calculus fit well with cognitive psychology's meta-theory.

Consequently, many psychologists have spent a great deal of time and effort in search of a "deductive component" (Wason & Johnson-Laird, 1972), or, in more current parlance, a logic "module." The theoretical burden they have placed on this proposed mental algorithm is staggering: It is supposed to be necessary for building virtually all the vast and complex knowledge structures that power human thought and behavior, from the most ubiquitous of social interactions to the most esoteric feat of modern technology.

1.2 What would a logic module be like?

The "doctrine of mental logic" (Johnson-Laird, 1982) is the view that the human mind includes innate algorithms instantiating the rules of inference of the propositional calculus -- a logic module. What properties can a logic module be expected to have?

Chomsky (1975), Marr & Nishihara (1978), and Fodor (1983) have taken the biological view (best summarized by Williams, 1966) that if a function is evolutionarily important, natural selection will produce a species-wide psychological mechanism with certain properties. Namely:

1. It will be specially designed to solve the evolutionary problem quickly, reliably, and efficiently. Consequently, it will instantiate mental architecture and rules of inference that will define the evolutionarily salient dimensions of the problem, and guide the organism toward an adaptively appropriate solution.

2. It will be domain specific. Only by limiting its scope of application can it be specially designed to solve the problem quickly and efficiently. I would add that it must have design features that make it sensitive to cues that indicate when the organism has encountered the domain for which the mechanism was designed. An algorithm that allows you to decide between fight or flight in the presence of a predator is useless unless it has features that let you determine what counts as a predator and when you are in the presence of one.

3. It will develop without explicit teaching or training. Exposure to the domain may be necessary to activate the mechanism or to allow it to fill in parameter values. But the rules that organize and process the stimuli are innately specified.

4. The inferences will be made automatically, without the application of "conscious effort" or deliberation. This is a consequence of its having to be fast and reliable (to remain reliable, the rules must be protected from the effects of deliberation -- they cannot be "isotropic" (Fodor 1983)).

Following this view, a logic module necessary for generating vast knowledge structures ought to have several properties:

Criterion A. It should instantiate procedures that reliably lead to valid deductions. Otherwise, it would not let you reject invalid hypotheses, and that is its proposed function.

Criterion B. It should be able to "recognize" hypotheses (in or out of consciousness), and upon recognizing them, correctly process them. This is because its domain is the universe of possible hypotheses.

Criterion C. It should process hypotheses quickly, automatically, and without "conscious attention." There are an infinite number of ways of carving the world into properties, and therefore an infinite number of relations between properties to serve as hypotheses; on average, an enormous number of hypotheses will have to be tested before the correct one is hit upon and the simplest generalization made (this point may be fatal to the entire learning-as-hypothesis-testing view). Therefore, processing must be quick and automatic.

Criterion D. It should develop without any special teaching. Adults rarely sit down and teach children the canons of formal logic, yet children learn things constantly. This means one of two things: either deduction is not necessary for most learning, or the logic module (or a "logic module acquisition device"!) is innate. If the logic module is necessary for learning, then it itself cannot be learned (Johnson-Laird, 1982). Hence, supporters of a hypothesis-testing view of learning are committed to an innateness position (whether they realize it or not).

Criterion E. With respect to the propositional content of the hypotheses it processes (what P and Q stand for in "If P then Q"), it should be content-independent. Because this mechanism is supposed to account for learning in all domains, the domain from which the propositional content of the hypothesis is taken should have no effect on how quickly the deduction is made or how likely it is to be valid.

The "doctrine of mental logic" was inspired primarily by criteria A and E. The propositional calculus is a system of rules for the derivation of valid inferences from propositions linked by logical connectives like and, not, and or. A logic module instantiating the propositional calculus would therefore satisfy criterion A, that the module instantiate procedures that reliably lead to valid deductions. In addition, conclusions derived via the propositional calculus are valid regardless of the specific content of the propositions involved. Its rules depend only on the truth values assigned to the propositions (whether the individual propositions are considered true or false) and on their position with respect to the logical connectives. For example, the "truth tables" associated with a conditional statement (If P then Q) and a biconditional statement (P if and only if Q) are:

Conditional

P	Q	If P then Q
T	T	T
T	F	F
F	T	T
F	F	T

Biconditional

P	Q	P iff Q
T	T	T
F	F	T
T	F	F
F	T	F

Thus, if P and Q are both considered true, then "If P then Q" is also considered true. Therefore, if P stands for "the sea is blue" and Q stands for "quantum physics is correct" -- and both these statements are considered true -- then the statement "If the sea is blue then quantum physics is correct" is also considered true. This property of the propositional calculus satisfies criterion E, that the logic module's rules of inference be content-independent.

1.3 Do humans have a logic module?

Above, I argued that if learning occurs through Popperian hypothesis testing, then humans must have a logic module -- a psychological mechanism with algorithms that allow people who have had no special training in logic to recognize hypotheses and deduce only their valid implications, quickly, reliably, and automatically. Research on deductive reasoning indicates that humans have no such ability (for reviews, see Wason & Johnson-Laird, 1972; Johnson-Laird, 1982). Because there is so little dissent on this point among psychologists who study logical reasoning, I will cite only a few illustrative examples, drawn from the literature on reasoning about conditional statements.

In reasoning about conditional statements, one can make two correct inferences and two fallacious inferences (to convince yourself, inspect the truth table in section 1.2):

Correct inferences		Fallacious inferences	
Modus ponens	Modus tollens	*Affirming the Consequent	*Denying the Antecedent
If P then Q	If P then Q	If P then Q	If P the Q
P	not-Q	Q	not-P
Therefore Q	Therefore not-P	Therefore P	Therefore not-Q

* These inferences are fallacious because a conditional does not claim that P is the only possible antecedent of Q. Consider a concrete, causal statement: "If it rains then the grass is wet." If it has not rained, the grass may or may not be wet -- perhaps I have been watering the lawn with my sprinkler. To conclude "it rained" (or "it did not rain") from the rule premise and the "grass is wet" premise is to "affirm the consequent". To conclude "the grass is dry" (or "the grass is wet") from the rule premise and the "it did not rain" premise is to "deny the antecedent". No valid inference can be drawn from these sets of premises.

Minimally, a logic module capable of evaluating conditional hypotheses should instantiate procedures that quickly and reliably accomplish modus ponens and modus tollens. Furthermore, it should be immune to the two fallacious inferences. The algorithms involved are simple and well-defined -- a microcomputer can easily be programmed to run them.

Moreover, the only fair way to test for a logic module is to use statements that express unfamiliar relations, such as "If an object is a triangle, then it is red", or "If there is an A on one side of the card, then there is a 3 on the other side". The logic module is supposed to be necessary for learning, that is, for the construction of new knowledge. If this is its purpose, then it should be good at handling unfamiliar relations. Furthermore, the use of relations drawn from unfamiliar domains provides a cleaner experimental design. It prevents subjects from simply "looking up facts" to answer the question: one need not engage in any reasoning process to decide that "All swans are orange" is false. Conditionals relating letters to numbers are good candidates because letters and numbers are familiar enough but relations between them are not. For lack of a better term, I will follow the literature and call such conditionals "abstract."

Prediction A: Valid deductions are made frequently and reliably.

Item. Shapiro (reported in Wason & Johnson-Laird, 1972, pp. 43-44) asked 20 subjects to evaluate the validity of abstract versions of the four inferences listed above. If humans have a logic module, her subjects should make few if any errors: they should judge the first two inferences valid and the last two inferences invalid. Errors should be randomly distributed among the four inferences. This task is very simple -- it does not even require subjects to generate conclusions themselves. All they have to do is correctly recognize inferences that have already been made as valid or invalid.

The error rate was reasonably low for modus ponens (5%), but the error rate was 52.5% for modus tollens, 20% for affirming the consequent, and 25% for denying the antecedent. Half the time subjects were judging a valid inference invalid, and a quarter of the time they were judging invalid inferences valid. Errors were not randomly distributed among the four conditions. The distribution of errors indicates that subjects find it particularly difficult to recognize the validity of modus tollens.

Item. In an experiment by Gibbs (reported in Wason & Johnson-Laird, 1972, p. 57-59) subjects had to generate deductions. On average, 44% of the problems requiring the use of modus ponens were done incorrectly, and 80% of those requiring modus tollens were done incorrectly. In both cases, incorrect inferences corresponded to committing the fallacy of affirming the consequent. Modus ponens was correctly used 2.8 times as often as modus tollens was.

Item. Mazzocco (reported in Legrenzi, 1970) found that subjects erroneously assume that "If P then Q" is equivalent to "If Q then P" when this makes a problem easier to "solve". Pollard & Evans (1980) found that subjects frequently view logically distinct conditionals as implying one another.

Item. Pollard & Evans (1981) found that subjects have a pronounced tendency to judge an inference valid when they agree with the conclusion, and invalid when they do not agree with the conclusion -- regardless of its true validity.

The claim that humans have a quick and reliable deductive component seems to fall before it takes its first step. Experimental results do not even support criterion A, that people be able to reliably make valid deductions. Although people have some measure of success in recognizing the validity of modus ponens, they are not good at using it to generate deductions. They are quite susceptible to making fallacious inferences, and they seem to lack a procedure corresponding to modus tollens almost entirely. The literature on logical reasoning is quite consistent on this point. According to Johnson-Laird (1982), "the doctrine of logical infallibility is either falsified by the results of some experiments on syllogistic reasoning or else empirically vacuous."

To save this perspective, one might argue that the logic module has a simpler design and a more specific function. Perhaps it does not have procedures for deriving deductive implications at all: perhaps it can only look for falsification. Nothing could be simpler to program. Consider any hypothesis of the form, "If P then Q." The truth table for the conditional shows that there is only one circumstance that can falsify this hypothesis: the co-occurrence of P and not-Q. A logic module capable only of falsification would scan all instances of P and all instances of not-Q. It would reject the hypothesis if any P was paired with a not-Q or if any not-Q was paired with a P.

The Wason selection task tests this prediction. Peter Wason was interested in Popper's view that the structure of science was hypothetico-deductive. The selection task allows one to see whether people really are falsificationists in testing hypotheses. In the selection task, a subject asked to test a hypothesis of the form "If P then Q" with respect to a universe of four cards representing possible pairings of P and not-P with Q and not-Q. Here is the original selection task (Wason, 1966):

Consider the following sentence:

"If a card has a vowel on one side then it has an even number on the other side."

It refers to these four cards:

```
.......    .......    .......    .......
:  E  :    :  4  :    :  K  :    :  7  :
:.....:    :.....:    :.....:    :.....:
```

Each card has a letter on one side and a number on the other side. Name those cards, and only those cards, which need to be turned over in order to determine decisively whether the sentence is true or false.

The cards were real cards, and an experimenter administered the task in person to one subject at a time.

The Wason selection task has a general solution. Turn over all cards displaying P (to see if they have a not-Q on the other side) and turn over all cards displaying not-Q (to see if they have a P on the other side). There is no point in turning over Q or not-P, because any value on the other side of these cards would be consistent with the hypothesis.

This provides a direct test of the modified view of the logic module's function. If the logic module is specialized for testing hypotheses through deductive falsification, then subjects should immediately realize that they must turn over the E card (P) and the 7 card (not-Q).

They do not. On average, only 4 to 10 percent of all subjects choose P and not-Q when confronted with an abstract hypothesis (Wason, 1983). The majority pick only P, or P and Q, as if they are trying to confirm the existence of a relation, rather than falsify a proposed relation. This result has been replicated many times, under a wide variety of conditions: with different abstract propositions standing in for P and Q, with variations in the linguistic format of the hypothesis, with variations in how the information is represented on the cards, with variations in the instructions (e.g., Wason, 1968; Wason, 1969a & b; Wason & Johnson-Laird, 1970; Wason & Shapiro, 1971; Goodwin & Wason, 1972; Wason & Golding, 1974).

Furthermore, it is very difficult to teach people the solution. A wide variety of "therapies" have been tried; they have resulted in little or no facilitation in falsification rates. For example:

1. For each of 24 sample cards, subjects were asked whether or not each card was consistent with the rule; they were given feedback about their answers. They were then asked to solve a selection task using the same rule (Wason & Shapiro, 1971).
2. For each of 24 sample cards, subjects were asked to imagine a value on the other side of the card that would falsify or verify the rule; they were given feedback about their answers. They were then asked to solve a selection task using the same rule (Wason & Shapiro, 1971).

3. Subjects were allowed to turn over the cards they had selected, asked whether each verified or falsified the rule, and corrected if wrong; then they were retested (Hughes, 1966).
4. A duplicate set of fully revealed cards were present for subjects to inspect (Goodwin & Wason, 1972).

Even professional logicians have been known to get the problem wrong! (Wason & Johnson-Laird, 1972, p. 179)

Subjects' performance on the Wason selection task is the most damning evidence against the learning-as-hypothesis-testing view. Confronted with a novel hypothesis, subjects do not try to falsify it. Yet this is a paradigmatic case in which they should use deductive falsification. This result falsifies the modified view of the logic module as specialized for spotting falsifying evidence. In addition, because modus ponens and modus tollens can be used to solve the selection task, this result, like the previously cited evidence, falsifies the original view of the logic module as instantiating deductive procedures to be used in falsifying hypotheses.

Beating a dead horse.

A logic module necessary for learning should meet four other criteria (B-E), but these are predicated on it fulfilling prediction A -- that people frequently and reliably make valid deductions. Prediction A has been shown to be false, so technically, the other predictions fall with it. Just to be thorough, however, I would like to briefly discuss each separately.

Prediction B: The logic module can "recognize" hypotheses, and upon recognizing them, process them.

Hypotheses about the world do not come in just one linguistic format. A logic module should be able to recognize and operate on the logical "deep structure" of a hypothesis, producing valid deductions regardless of its linguistic format. The amount of time the conversion to deep structure takes may differ with linguistic format, not the validity of the deductions made.

This is not the case. A number of studies show that (1) different linguistic formats of the same hypothesis differ in how likely they are to elicit a valid deduction, and (2) a linguistic format that facilitates deduction for one problem may impede deduction in another (e.g., Van Duyne, 1974; Roberge, 1978, 1982; Bracewell & Hidi, 1974). Subjects in these studies had no time constraints, so differences in performance can be accounted for only by differences in linguistic format.

Prediction C: Valid deductions are made quickly, automatically, and without conscious attention.

In the experiments cited under prediction A, subjects were permitted to devote all the time and conscious attention to the problem that they wanted, yet they still did not make valid deductions. Clearly they do not make valid deductions quickly, automatically, and without conscious attention.

Prediction D: The logic module develops without any special teaching.

Again, the evidence cited for prediction A shows that people do not reliably make valid deductions without special training. Indeed, it is not clear that they reliably make valid deductions even with special training. As the therapy experiments showed, performance on the Wason selection task proved relatively impervious to special training techniques, and even professional logicians find it difficult.

Prediction E: The logic module is content-independent.

Wason's first selection tasks used hypotheses that expressed abstract relations, usually involving letters and numbers. Performance was uniformly poor. However, a number of experiments in the early 1970's reversed this result (Wason & Shapiro, 1971; Johnson-Laird, Legrenzi & Legrenzi, 1972; Bracewell & Hidi, 1974; Gilhooly & Falconer, 1974). These experiments suggested that if the content of the rule being tested expresses a "familiar," "realistic," or "thematic" relation, subjects do reason logically on the selection task. This enhancement of logical performance with familiar materials is known as the "content effect" or the "thematic materials" effect on the Wason selection task.

Initially, researchers thought that the familiarity or realism of thematic content somehow facilitates the use of deductive logic (Wason & Shapiro, 1971; Johnson-Laird, Legrenzi & Legrenzi, 1972). The problem with this explanation is that the phenomenon is quite difficult to replicate. Some familiar content seems to facilitate the use of deductive logic; other familiar content does not (e.g., Van Duyne, 1976; Manktelow & Evans, 1979; Griggs & Cox, 1982; Cox & Griggs, 1982; Reich & Ruth, 1982; Yachanin & Tweney, 1982; Griggs & Cox, 1983). In addition, the same familiar content seems to facilitate logic at some testing locations, but not at others (e.g., Golding, 1981;

Griggs & Cox, 1982; Yachanin & Tweney, 1982). This should not happen if familiar content simply activates a logic module.

Whatever the explanation, the cognitive processes that govern reasoning about logical conditionals in the Wason selection task are clearly not content-independent.

The hypothesis that humans have the sort of logic module necessary for Popperian-style everyday learning faltered before taking its first step. Not one of the five defining criteria of a logic module is fulfilled by the results of experiments on logical reasoning.

This raises some serious questions: If people are not using deductive rules to reason about conditional statements, then what rules are they using? And if people are not learning via Popperian hypothesis testing, then how are they learning?

The discovery of the content effect on the Wason selection task raises the possibility that reasoning about logical conditionals is governed by content-dependent cognitive processes. Indeed, after years spent researching this effect, Wason and Johnson-Laird commented that the conditional "is not a creature of constant hue, but chameleon-like, takes on the colour of its surroundings: *its meaning is determined to some extent by the very propositions it connects*" (1972, p.92, italics theirs). They say that the principles governing the "cohesion of discourse" probably hold the key to its many meanings, and that "the nature of these principles is little understood -- they probably involve more than purely linguistic factors." The investigation of Darwinian algorithms presented in this thesis is a preliminary enquiry into what "more" they involve.