An End to the Controversy? A Reply to Rips

Philip N. Johnson-Laird*
Department of Psychology
Princeton University
Princeton, NJ 08544, U.S.A.
phil@clarity.princeton.edu

There is a controversy about the psychology of reasoning. Its crux is:

> When logically-untrained individuals reason, do they rely on syntax or semantics? That is, do they rely on formal proofs or mental models?

Lance J. Rips's *The Psychology of Proof* (Rips 1994) defended formal proofs, which his PSYCOP theory derives from rules of inference like those of a logical calculus. In my review (Johnson-Laird 1997), I congratulated him on having formulated the best available formal theory, but I defended mental models. They are the end product of perception and of understanding discourse, and they can be used for reasoning by testing whether any models of the premises are counterexamples to conclusions. Rips's reply to my review (Rips 1997) clarifies the goals of his theory, brings out some further points in its favor, responds to my criticisms, and attacks the mental-model theory. Yet, despite our negative assessments of one another's theories, Rips and I have much in common. We both believe that deduction is a major mental ability, that cognitive scientists should implement their theories in computer programs, and that the controversy will be resolved by empirical observations. What is needed, in fact, is a robust and surprising phenomenon that one theory predicts and that the other theory neither predicts nor accommodates. This note describes such a phenomenon, but it begins with Rips's reply.

1 Rips's Reply to the Case against PSYCOP

My review described some possible flaws in PSYCOP judged on its own terms as a formal theory of deduction:

Individuals are supposed to recognize PSYCOP's formal rules as intuitively sound. Yet it contains rules that most participants in Rips's own experiments have difficulty in using, e.g., the rule for introducing "or". Likewise, the rule for using *modus ponens* in a backward chain of inference has nine separate clauses in its formulation. Rips counters that it is hard to see how cognitive psychology

* I am most grateful to Ruth Byrne for her work in developing the model theory, and to Fabien Savary for his help in exploring illusory inferences. Thanks also to Lance Rips for his comments on an earlier draft of this reply, and for conducting himself at all times according to the Marquess of Queensbury rules. And thanks to the Department of Psychology, New York University, which made me a most welcome guest during the writing of this paper. Sadly, my host, the late Martin Braine, was too ill to continue our long-standing discussions. The research was supported in part by ARPA (CAETI) contracts N66001-94-C-6045 and N66001-95-C-8605.

could progress if it were to limit nonconscious processes to simple routines. Some psychologists – connectionists, perhaps – might beg to differ. In any case, the status of PSYCOP's rules as "intuitively sound" is not clear-cut.

PSYCOP is "incomplete"; i.e., it cannot prove certain valid inferences, and, as Jon Barwise (1993) has shown, incompleteness vitiates psychological theories. When PSYCOP fails to prove a conclusion, it may fail because it cannot derive a proof or because the conclusion is invalid. Rips replies that people *are* often uncertain whether their inability to find a proof (on a math test, for example) means that no proof exists or that they have not been able to find it. True; but consider the following sort of argument:

If there is a short, then the battery is dead.
The battery isn't dead.
Therefore, it will rain today.

PSYCOP tries to find a derivation, fails, and gives up. If people were like PSYCOP, they would never *know* that the inference was invalid. Ironically, Rips (1994) showed that PSYCOP could be made complete by adding, in effect, the following rule:

$$\frac{\text{NOT(IF P THEN Q)}}{\text{P AND (NOT Q)}}$$

Why not add this rule to PSYCOP? Rips demurs, because he doubts that it is intuitively obvious. Yet several experimenters have asked people to construct instances that would falsify a conditional of the form: if p then q. They reliably respond with cases of: p and not q (see, e.g., Oaksford and Stenning 1992). Granted that Rips's Goal Number 1 is: "Construct a First-Order Deduction System", it is odd that he settles for incompleteness.

Unlike PSYCOP, people *are* able to construct counterexamples, i.e., to imagine situations in which the premises are true but an invalid conclusion is false. Rips replies: "There is no doubt, for example, that people can learn devices like Euler circles or Venn diagrams and can use them to test syllogisms by searching for counterexamples." What he has in mind is not PSYCOP itself, which cannot search for counterexamples, but his Deduction-System Hypothesis. According to this hypothesis, PSYCOP is also a theory of cognitive architecture, and it can be used as a language in which to write programs that simulate skills that PSYCOP itself lacks. The problem with this defense is that it renders the theory irrefutable. Suppose, for example, that psychologists discover that reasoning depends on a computable procedure, X, that is not part of PSYCOP. The discovery does not jeopardize PSYCOP, because it can be used to write a program that carries out X. In this way, one always avoids the unhappy conclusion that PSYCOP is wrong. Indeed, X could be the theory of mental models!

In PSYCOP, the rules that make suppositions can be used only in backward chains of inference (see the Tables of rules in my review). I argued that this step is in the wrong direction, because Rips had formerly allowed suppositions to

be made in forward chains too. Rips responds (note 9): "... PSYCOP retains the ability to reason both forward and backward from assumptions." My point, however, did not concern reasoning from assumptions once they have been made, but the very act of making an assumption. PSYCOP can *make* an assumption only in a backward chain, whereas people can make assumptions without a possible conclusion in mind.

In general, the model theory makes parameter-free predictions, whereas PSYCOP uses parameter estimates to accommodate data *a posteriori*. My review cited five phenomena that only the model theory predicts:

1. Reasoning with conjunctions is easier than reasoning with conditionals, which in turn is easier than reasoning with disjunctions. Rips counters that an inference of the form:

Not both p and q.
p.
Therefore, not q.

is more difficult than an argument of the form:

p or q.
not p.
Therefore, q.

But my claim was about *unnegated* connectives. The model theory predicts that negating a conjunction should increase its difficulty. It increases the number of models. A conjunction, p and q, calls for a single model:

$$p \quad q$$

whereas a negated conjunction – not both p and q – calls for three models, shown here on separate lines in a fully explicit form (where "$-$" signifies negation):

$$\begin{array}{rr} p & -q \\ -p & q \\ -p & -q \end{array}$$

One minor misunderstanding: Rips writes, "... Johnson-Laird's mental models don't include information about either scope or variables." In fact, mental models are constructed from propositional representations that include information about both scope and variables (Johnson-Laird and Byrne 1991, Ch. 9), and negation should be harder to understand when it has more than one proposition within its scope. Hence, the inference from the negated conjunction should be more difficult than the inference from the disjunction. Rips's putative counterexample is thus a phenomenon that the model theory predicts but that the rule theory merely accommodates.

2. *Modus ponens* can be suppressed by certain additional premises (Byrne 1989). Rips points out that the cause is hotly debated, and cites Guy Politzer and Martin Braine (1991), who argue that the additional premise casts doubt

on the truth of the conditional. Ruth Byrne (1991), however, has demonstrated suppression without affecting the believability of the conditional.

3. Reasoning with exclusive disjunctions is easier than reasoning with inclusive disjunctions; e.g.:

p or else *q*, but not both.
not *p*.
Therefore, *q*.

is reliably easier than:

p or *q*, or both.
not *p*.
Therefore, *q*.

Rips says that there is no difference in some early studies. Maybe. But all the studies that detect a *reliable* difference corroborate the model theory's prediction (e.g., Roberge 1976; Johnson-Laird, Byrne, and Schaeken 1992; Bauer and Johnson-Laird 1993).

4. Certain diagrams improve reasoning with disjunctions (Bauer and Johnson-Laird 1993). Rips counters that training with diagrams may not transfer to other sorts of reasoning. But training is not at issue, because there was none in our studies. Rips also argues that the diagrams that people use in solving syllogisms tend to be variations on the Euler circles that they learned in school. I agree. However, in a recent unpublished study, Fabien Savary and I videotaped individuals as they carried out sentential reasoning. They often used their own idiosyncratic diagrams. Given the premises:

Either there is a grey marble or else there is a brown marble, but not both.
There is a brown marble if and only if there is a white marble.
Either there is a white marble or else there is a blue marble, but not both.

one person, for example, drew a diagram in which each row represents a possible state of affairs:

blue	grey
white	brown

Another person drew a vertical line and then added the colors in the following arrangement:

	white
grey	brown
blue	

It has not escaped our notice that such diagrams are isomorphic to mental models of the premises.

5. PSYCOP provides no account of systematic errors. Rips replies that errors may occur at any point in the process of reasoning. I agree. But, as we will now see, certain systematic errors confirm the model theory and disconfirm formal theories.

2 The Case for Mental Models

Both PSYCOP and the model theory apply to deductions based on sentential connectives and quantifiers. Unlike PSYCOP, however, the model theory integrates deduction, probabilistic reasoning, and modal reasoning. A conclusion is *necessary* if it holds in all the models of the premises, a conclusion is *probable* if it holds in most of the models of the premises, and a conclusion is *possible* if it holds in at least some model of the premises. The model theory also applies to such quantifiers as "more than half" that are outside the first-order calculus (see Johnson-Laird 1983). And it applies to the informal arguments in scientific articles, newspaper editorials, and legal proceedings. A striking fact about such arguments is that they are so remote from formal proofs that many theorists argue that logic plays no part in their construction or evaluation (see, e.g., Toulmin 1958). Such claims throw out the model baby with the unruly – or at least unrule-like – bath water. That is to say, informal arguments are constructed with a view to validity in terms of models, not formal proofs (Shaw 1996). In sum, the model theory has a wider range of application than PSYCOP.

What matters in resolving a controversy, however, is not the range of theories, but a crucial phenomenon. The fundamental *representational* assumption of the model theory is that individuals reduce the load on working memory by representing explicitly only those cases that are true (Johnson-Laird and Byrne 1991, Ch. 3). This assumption is subtle, as the following example illustrates:

There is a king in the hand or else there is an ace in the hand, but not both.

The assertion calls for just two models:

king

ace

These models correspond to the true cases, but, most importantly, they represent only their true parts. Thus, the first model does not represent explicitly that it is false that there is an ace in this case, and the second model does not represent explicitly that it is false that there is a king in this case. Individuals should make "mental footnotes" to allow them to reconstruct this information, but these footnotes are easily forgotten with more complex assertions.

The fundamental assumption leads to a surprising prediction that we discovered by accident in the output of a computer program implementing the model theory. There should be illusory inferences in which individuals go seriously wrong from failing to represent what is false (Johnson-Laird and Savary 1996a). Here is an example akin to one in my review:

If there is a king in the hand, then there is an ace,
or else if there isn't a king in the hand, then there is an ace.
There is a king in the hand.

Given these premises, almost everyone draws the conclusion:

There is an ace in the hand.

Savary and I, together with various colleagues, have observed it experimentally (Johnson-Laird and Savary 1996b); we have observed it anecdotally – only one person among the many distinguished cognitive scientists to whom we have given the problem got the right answer; and we have observed it in public lectures – several hundred individuals from Stockholm to Seattle have drawn it, and no one has ever offered any other conclusion. Yet, as I explained in my review, it is wrong. People fail to consider that when one conditional is true, the other conditional is false.

Of course, one result is hardly crucial. It is open to several alternative explanations. Thus, Rips in his reply speculates that perhaps reasoners paraphrase the disjunction as:

If there is either a king or no king in the hand, then there is an ace.

We now know, however, that this hypothesis is not the correct explanation of illusory inferences. It seems less plausible and more *ad hoc* when the disjunction is described in the following terms, which still give rise to the illusion (Johnson-Laird and Savary 1996b):

One of these assertions is true and one of them is false:

If there is a king in his hand, then there is an ace in his hand.
If there is not a king in his hand, then there is an ace in his hand.

And Rips's hypothesis clearly cannot explain those illusions – again, predicted by the model theory – that do not depend on conditionals, e.g.:

Only one of the following two assertions is true:

Albert is here or Betty is here, or both.
Charlie is here or Betty is here, or both.

This assertion is definitely true:

Albert isn't here and Charlie isn't here.

These premises yield the illusion that Betty is here. With control problems, as the model theory predicts, the failure to represent false cases does not lead to systematic errors.

Illusory inferences contravene PSYCOP and the other formal rule theories. These accounts all rely solely on valid rules, and so they cannot account for inferences in which most people draw one and the same wrong conclusion. Rips, however, argues that the illusions are equally problematic for the model theory. He writes: "they devastate the theory presented in Johnson-Laird, Byrne, and Schaeken 1994", on the grounds that we argued that erroneous conclusions should be consistent with the premises. But it is a mistake to confuse the theorists with their theory. The theory in Johnson-Laird et al. 1992 and 1994 is correct: It predicts the illusory inferences. Its authors, however, were wrong about their own theory, because they had yet to discover that it predicted the illusions. Erroneous

conclusions do indeed correspond to initial models of premises, which are usually consistent with the premises. But, just sometimes – in the case of the illusions – they are inconsistent with the premises. In short, illusory inferences devastate the theorists, not their theory.

Here is how matters stand on the controversy about reasoning: The semantic theory predicts a surprising phenomenon; the syntactic theories neither predict nor accommodate it. Our results suggest that the illusions are varied and robust. But, as Rips has rightly reminded us, much of the evidence has yet to be corroborated. So, is this the end of the controversy as we know and love it? Certainly, it seems to be the beginning of the end.

References

1. Barwise, Jon (1993), 'Everyday Reasoning and Logical Inference', *Behavioral and Brain Sciences* 16, pp. 337–338.

2. Bauer, Malcolm I., and Johnson-Laird, P. N. (1993), 'How Diagrams Can Improve Reasoning', *Psychological Science* 4, pp. 372–378.

3. Byrne, Ruth M. J. (1989), 'Suppressing Valid Inferences with Conditionals', *Cognition* 31, pp. 61–83.

4. Byrne, Ruth M. J. (1991), 'Can Valid Inferences Be Suppressed?', *Cognition* 39, pp. 71–78.

5. Johnson-Laird, P. N. (1983), *Mental Models*, Cambridge, MA: Harvard University Press; Cambridge, UK: Cambridge University Press.

6. Johnson-Laird, P. N. (1997), 'Rules and Illusions: A Critical Study of Rips's *The Psychology of Proof*', *Minds and Machines* 00, pp. 000–000.

7. Johnson-Laird, P. N., and Byrne, Ruth M. J. (1991), *Deduction*, Hillsdale, NJ: Lawrence Erlbaum Associates.

8. Johnson-Laird, P. N.; Byrne, Ruth M. J.; and Schaeken, Walter S. (1992), 'Propositional Reasoning by Model', *Psychological Review* 99, pp. 418–439.

9. Johnson-Laird, P. N.; Byrne, Ruth M. J.; and Schaeken, Walter S. (1994), 'Why Models Rather than Rules Give a Better Account of Propositional Reasoning: A Reply to Bonatti, and to O'Brien, Braine, and Yang', *Psychological Review* 101, pp. 734–739.

10. Johnson-Laird, P. N., and Savary, Fabien (1996a), 'Illusory Inferences about Probabilities', *Acta Psychologica* 93: 69-90.

11. Johnson-Laird, P. N., and Savary, Fabien (1996b), 'Truth and Illusion in Reasoning', under submission.

12. Oaksford, Michael, and Stenning, Keith (1992), 'Reasoning with Conditionals Containing Negated Constituents', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, pp. 835–854.

13. Politzer, Guy, and Braine, Martin D. S. (1991), 'Responses to Inconsistent Premises Cannot Count as Suppression of Valid Inferences', *Cognition* 38, pp. 103–108.

14. Rips, Lance J. (1994), *The Psychology of Proof: Deductive Reasoning in Human Thinking*, Cambridge, MA: MIT Press.

15. Rips, Lance J. (1997), 'Goals for a Theory of Deduction: Reply to Johnson-Laird', *Minds and Machines* 00, pp. 000–000.

16. Roberge, J. J. (1976), 'The Effect of Negation on Adults' Disjunctive Reasoning Abilities', *Journal of General Psychology* 91, pp. 23–28.

17. Shaw, Victoria F. (1996), 'The Cognitive Processes in Informal Reasoning', *Thinking and Reasoning* 2, pp. 51–80.

18. Toulmin, Stephen E. (1958), *The Uses of Argument*, Cambridge, UK: Cambridge University Press.