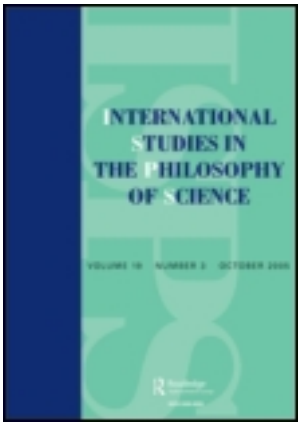

International Studies in the Philosophy of Science

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cisp20

Reply to the commentators on a model theory of induction

Philip N. Johnson-Laird [a]

[a] Department of Psychology, Princeton University, New Jersey, 08544, USA

Version of record first published: 09 Jun 2008.

To cite this article: Philip N. Johnson-Laird (1994): Reply to the commentators on a model theory of induction, International Studies in the Philosophy of Science, 8:1, 73-96

To link to this article: http://dx.doi.org/10.1080/02698599408573485

Reply to the commentators on a model theory of induction

PHILIP N. JOHNSON-LAIRD
Department of Psychology, Princeton University, New Jersey 08544, USA

Introduction

The target article had three goals: to distinguish between induction, deduction and creation; to outline a taxonomy of sorts of induction; and to develop a new theory of inductive reasoning based on mental models. The commentators have raised a rich and stimulating set of issues, and I am grateful to them for forcing me both to reconsider certain matters and also to think for the first time about others. They point out omissions and weaknesses in the mental-model theory, but the aim of this reply is to show that their objections are not decisive. It will establish that some of their worries are groundless or due to misunderstandings, some can be accommodated by the existing theory, and some call for extensions of the theory that the paper will outline. The model theory is still viable and offers a plausible explanation of inductive reasoning.

The reply will finesse at least two issues. The first concerns concepts. Newstead asks how concepts relate to one another, and how the model theory deal with 'fuzzy' concepts and prototypes. Garnham similarly asks how individuals construct new concepts, or new versions of old concepts in a scientific revolution. Elsewhere, I have recently tried to deal with these matters (Johnson-Laird, 1993), and I spare readers a repeat of my efforts here. The second matter is Bara's claim that the key to understanding mental processes is the study of their development in childhood. The claim is plausible. But intellectual development may be a process of conceptual change rather than a passage from one sort of thinking to another in a sequence of distinct stages (see e.g. Carey, 1985). If so, developmental studies may not be the key to understanding mental processes. In any case, the innate endowment required for inductive learning and a sketch of conceptual development can also be found in Johnson-Laird (1993).

The broad plan of the reply is to deal in the first part with commentators' reactions to the proposed theoretical distinction between induction, deduction, and creation, and to the use of semantic information in setting up a taxonomy of thinking. The second part of the paper takes up the single most blatant, though deliberate, omission in the target article—the model theory's treatment of probability and the strength of inductions. The third part of the paper considers the relations between models and rules (or schemas) for induction. The paper finally draws some general conclusions about the role of models in thinking.

1. The taxonomy of thinking

The target article proposed a taxonomy of thinking based on the concept of semantic information. The theory applies only to thoughts with a propositional content, and the taxonomy depends on the five possible semantic relations between premises and conclusion:

(1) The premises and conclusion rule out exactly the same states of affairs.
(2) The conclusion rules out fewer states of affairs than the premises, i.e. it throws some semantic information away.
(3) The premises and conclusion rule out disjoint states of affairs.
(4) The premises and conclusion rule out overlapping states of affairs.
(5) The conclusion goes beyond the premises to rule out some additional state of affairs over and above what they rule out.

The first two of these relations are valid deductions, i.e. cases where if the premises are true then the conclusion must be true too. As Smith observes, however, people do indeed boggle at deductions that merely throw semantic information away, e.g. an inference of the form:

A.
∴ A or B, or both.

Such steps occur as part of other more general deductions that maintain information and also in making more specific an inductive generalization, e.g. an inference of the form:

If C then A.
∴ If C then (A or B, or both).

The third relation occurs in the step from premises to a conclusion that contradicts them. The fourth possibility is a creative relation in which some premises are abandoned and other new information is added. The fifth possibility is truly inductive, i.e. it embraces all the traditional cases of induction such as the generalization from a finite set of observations to a universal claim.

Several commentators raised objections or queries about the taxonomy and semantic information. The semantic information conveyed by a proposition is defined by the states of affairs that it rules out as impossible. A measure of semantic information can be formulated according to the following principle:

$$I(A) = 1 - p(A).$$

where I(A) is the semantic information in A, and p(A) is the probability that A is true (see Johnson-Laird, 1983, p. 34 *et seq.*). As this definition shows, there is a close relation between the taxonomy and probabilistic reasoning (more on this point in Part 2).

Green clarifies the taxonomy's implications by drawing a distinction between semantic information and psychological information. Semantic information depends solely on the states of affairs that propositions rule out, and thus deductions do not increase semantic information. Yet psychologically speaking, the conclusion of a deduction may be psychologically informative by making explicit what is merely implicit in the premises. Explicit information is immediately available for further processing. Implicit information is not so available: it can be made explicit but the process takes time and effort. Green also argues that both deduction and induction can be creative. The issue

here is the nature of the processes yielding the conclusion, and whether it would be revealing to characterize it as "creative". On the whole, processes that merely make explicit what was implicit are unlikely to satisfy the criteria of creativity (see Johnson-Laird, 1993, Ch. 3). The discovery of a deductive proof, however, may be highly creative: such a discovery does not itself depend on purely deductive steps, but may also call for inductions and creative conjectures. A mental process that makes an inductive step could arguably be creative because it adds new information, but since it changes nothing in its starting point, one might prefer to reserve the notion of creativity for more radical changes in conceptual content.

Over and Manktelow suggest that the analysis in terms of semantic information is too rigid. They argue that the definition of induction as the increase of semantic information is too narrow: even if the premises are held with certainty, induction should not utterly eliminate models, because it can be probabilistic. As we have seen, however, semantic information is more than merely compatible with notions of probability: its measure is defined in terms of probability. Over and Manktelow also remark that too strict a conception of semantic information leads to difficulties in the case where premises and conclusion rule out disjoint situations (the third relation in the list above). They claim, contrary to the target article, that there is a rational process of thought that exemplifies this relation: a pragmatic inference from a denial, as in the inference from an individual A's assertion:

There is no reason to think I'm a crook.

to the conclusion:

There is a reason to think A is a crook.

In the right context, as they say, the inference is non-demonstrative but rational. However, it is not an instance of premise and conclusion ruling out disjoint states of affairs. The inference does not have the form:

p.
∴ not p.

but rather the following form:

A asserts not p.
∴ p.

Both the premise and the conclusion could be true (as many hold about Nixon's celebrated remark, "I am not a crook"). What *would* be an inference exemplifying the third relation is:

There is no reason to think that A is a crook.
∴ There is a reason to think that A is a crook.

Immediate inferences of this sort, as the target article argued, do seem unlikely to occur.

Of all the commentators, Mosconi expressed the deepest misgivings about the taxonomy, and it may be instructive to consider them in detail. He begins with a point that is well-taken: everyday language has no terms for deduction and induction, and he asks why. The answer according to the model theory is because they are very similar mental processes—the only essential difference is that induction adds information to models whereas deduction maintains it. He points out that the taxonomy appears to reduce forms of thought other than deduction and induction to a residual class. That is

certainly not the intention: in fact, creative thinking is most important, but it also happens to be the variety of thought about which least is known. Mosconi also points out that creativity need not depend on premises. That, too, is true (cf. musical creativity in which the starting point is not a set of premises but a repertoire of elements that can be used to create rhythms, melodies, chord sequences, and so forth, see Johnson-Laird, 1993). What must be emphasized is that the analysis in the target paper concerns only thinking that concerns propositional content. But perhaps Mosconi has a more subtle point to make. He writes:

> In situations of problem-solving, the solution (which is often only a 'conclusion' in the sense that it is the final act of a process) may derive from, or fundamentally consist of a reinterpretation or transformation of the initially given elements or the terms of the problem; this is not the case in deduction or induction, because this would simply mean another deduction or induction.

The passage seems to imply that an analysis of problem solving in terms of semantic information is inappropriate, because the notion does not apply to the reinterpretation or transformation of information given in the statement of a problem. The claim needs arguing, because problems in cryptarithmetic, mechanics, and the repertory of experimental psychology (Tower of Hanoi, missionaries and cannibals, etc.) are all perfectly amenable to an analysis in terms of semantic information. Indeed, this insight underlies much of the development of AI approaches to problem solving as a form of deductive theorem-proving from PLANNER onwards (see e.g. Newell, 1990). The target article claims that each step in solving a problem with a propositional content falls into one of the five categories. The claim may be wrong, but its refutation calls for a demonstration that a step in solving a problem falls outside the five categories.

Mosconi offers the following view of deduction:

> In reasoning (and we can take the case of the deductive process), the premises are made with the aim of being able to draw a conclusion from them. For this to be reasonably possible, each premise must bear an independent (or more properly, different) item of information. The premises must be informative both in themselves and when they are considered in relation to each other. In other words, the propositions taken as premises must be synchronously acceptable.

He then argues that these conditions are not fulfilled by the target article's example of an individual who knows:

> The battery is dead or the voltmeter is faulty, or both.

who then tests the voltmeter and discovers that it is not faulty, and so draws the valid conclusion:

> The voltmeter is not faulty and the battery is dead.

Mosconi's claim is that the premises must be informative in themselves and when they are considered in relation to one another. However, he owes us an account of what he means by 'informative'. My best construal is as follows:

(1) The premises should not be inconsistent taken either individually or together.
(2) No premise should follow validly from others.

But this claim is psychologically unrealistic. You do not always have the opportunity to

choose your premises, and you are sometimes confronted with premises that are inconsistent. You can argue from premises that you do not believe (as in demonstrating their inconsistency), or from premises that are purely hypothetical. And if you were to try to abide by Mosconi's account, your premises should be synchronously acceptable, but you cannot know whether a set of premises is synchronously acceptable unless you have tested their acceptability, and this test usually calls for reasoning: you have to reason before you can reason—a desideratum that leads to a slippery recursive slope that may never 'bottom out', in which reasoning will be indefinitely postponed by the further need to reason about what premises to reason from. A simpler view, in contrast, is that the propositions which individuals take as premises do *not* have to be synchronously acceptable.

In his analysis of the deduction about the battery, Mosconi writes:

> Once we know that 'the voltmeter is not faulty', we can no longer say that 'the battery is dead or the voltmeter is faulty' ..., and therefore we cannot use these two propositions as premises. The second item of information (that the voltmeter is not faulty) eliminates one term in the disjunction, which can no longer be said in synchrony with the new proposition.

But, how do we know that the second item of information eliminates one term in the disjunction? The answer, of course, is that we deduce this consequence. The model theory proposes that the step depends on combining the models of the disjunctive premise with the model of the categorical premise to yield a single resulting model from which the conclusion can be read off:

The battery is dead.

The process of deduction allows us to reach a useful conclusion that makes explicit a state of affairs that is only implicit in the premises. We therefore need to distinguish between the basis for deductions and the conventions for asserting propositions. On the one hand, individuals often make inferences of the form:

p or q, or both.
not p.
∴ q.

which violate Mosconi's prescription. 'Trouble-shooting' and diagnosis of faults are replete with such cases. On the other hand, Mosconi is right that once one knows a proposition of the form:

not p.

it would throw semantic information away and thus violate Gricean conventions of discourse to assert:

p or q, or both.

Mosconi also has difficulties with the example of a specific induction:

The battery is dead or the voltmeter is faulty.
The voltmeter is faulty.
∴ The battery is not dead.

He finds it difficult to imagine that anyone would make this inference without having any new information. It is not clear what information he thinks is required, because he

is evidently happy to make an analogous leap from specific observations to a general conclusion.

Analogy, abduction, and what the taxonomy may overlook

When one examines the conventional chapter-headings of a text on thinking, as Garnham remarks, they reflect the traditional ways in which research has been done rather than underlying theoretical divisions. The taxonomy proposed in the target article is supposed to reflect an underlying theory at the molecular level. Thus, a person carrying out a task at the molar level, such as solving a problem, may carry out a sequence of steps, some of which are deductions, some inductions, some creations. The danger with a theoretically-driven taxonomy is that it may overlook some 'species' of thought, which would cast doubt not only on the taxonomy but also on the underlying theory. Green refers to two species of thought that are not mentioned in the target article—abduction and analogy—and so we should consider whether they can be accommodated within the taxonomy.

Abduction has a slightly confused history, but most authors, following Peirce (1958), use the term to refer to the process of generating explanations. An example is the step, say, from observing bees with a particularly potent sting to the hypothesis that the cause is a change in the bee's genetic structure—a step that may be an induction intended to explain an observed property. Green argues that abduction is based on what is possible whereas induction is based on what is probable, and that abduction supplies premises from background knowledge rather than from what is observed. These claims may be correct—certainly, the key step in abduction is to provide a hypothesis explaining an observation, whereas an induction may merely generalize an observation. In both cases, however, there is an inferential step that increases semantic information, and it is unclear that abduction has to be based on what is possible and induction on what is probable. In either case, reasoners may prefer the probable to the possible.

Analogy is more problematic for the proposed taxonomy. It has long been distinguished as a special sort of thinking, and to subsume it under the heading of induction appears to abandon its special status. However, Green suggests that analogical thinking can be explained in terms of mental models in at least two ways. The traditional way is to suppose that a reasoner recovers a source model and then transfers some structural relations from it to the target model in order to solve a problem (see e.g. Gentner, 1983). An alternative way, which is due to Imre Schlesinger, is to combine induction and deduction, e.g.:

	It is wrong to let people suffer.	[premise]
∴	It is wrong to let living creatures suffer.	[by induction]
	Animals are living creatures.	[premise]
∴	It is wrong to let animals suffer.	[by deduction]

Both approaches are feasible, though the second appears to be based on simple conceptual relations and so is unlikely to account for more profound scientific analogies, such as Bohr's explanation of atoms by analogy with the solar system.

Semantic information and quantifiers

Garnham asks how the notion of semantic information (and thus the elimination of models) applies to quantified domains. The theory has been worked out in some detail

but this section will sketch only the main results. Where quantifiers range over infinite sets of individuals, it is impossible to calculate semantic information on the basis of cardinalities, but it is possible to ascertain a partial rank order of generalization: one assertion is a generalization of another if it eliminates certain states of affairs over and above those eliminated by the original assertion. In the monadic predicate calculus, there are two generalization hierarchies in which each successive assertion is a generalization of the previous one:

$(\exists x)(Gx \vee Hx)$: There are some x that are G or H.
$(\exists x)(Hx)$: There are some x that are H.
$(\exists x)(Gx \wedge Hx)$: There are some x that are G and H.

and

$(\forall x)(Gx \vee Hx)$: Any x is G or H.
$(\forall x)(Hx)$: Any x is H.
$(\forall x)(Gx \wedge Hx)$: Any x is G and H.

The inverse of these hierarchies hold for the negations of these assertions.

There is also a hierarchy of generalization for assertions containing multiple quantifiers. Suppose, for instance, that their is a finite domain of discourse consisting of two disjoint sets of entities: a set of m entities with property D (e.g. 'drugs') and a set of n entities with property A (e.g. 'ailments'), where m may equal n, and that there is a binary relation, C, (e.g. 'cures') that can hold between ordered pairs of entities drawn from the two sets (e.g. drug i cures ailment j). The apparatus of restricted quantification enables us to write $(\exists_D x)$ in order to signify that there is an x with property D. The semantically weakest assertion about the domain is:

$(\exists_D x)(\exists_A y)(Cxy)$: Some drug cures some ailment

The number of possible states of affairs in this finite domain is equal to 2^{mn}, and hence the informativeness, I, of this assertion equals 2^{-mn} assuming that each state of affairs is equally possible. The only state of affairs that the assertion rules out is the one where no drugs cure any ailments. Universal generalization strengthens the assertion, and there are two possible generalizations depending on which existential quantifier is generalized:

$(\forall_D x)(\exists_A y)(Cxy)$: Every drug cures some ailment,
where $I = 1 - (2^n - 1)^m \ (2^{-mn})$

and

$(\forall_A y)(\exists_D x)(Cxy)$: Every ailment is cured by some drug,
where $I = 1 - (2^m - 1)^n \ (2^{-mn})$

The informativeness of these two assertions depends on the sizes of the two sets. If, say, there ae 3 drugs ($m = 3$) and 5 ailments ($n = 5$), the informativeness of the first assertion is about 0.09, whereas the informativeness of the second assertion is about 0.49. Where there are few drugs and many ailments, it is thus more informative to discover that every ailment is cured by a drug than to discover that every drug cures some ailment—from a medical standpoint, the former may also be more important too.

The next step in generalization can be brought about by shifting an existential quantifier from within the scope of a universal quantifier to immediately outside it. (This operation, as the target article mentioned, does not appear to have been exploited by any AI programs.) The first of the previous cases thus becomes:

$(\exists_A y)(\forall_D x)(Cxy)$: Some ailment is cured by every drug, where $I = (2^m - 1)^n\ (2^{-mn})$, i.e. about 0.51 in the present case.

The assertion rules out all of the states of affairs that the earlier assertion eliminates together with some additional states of affairs. The second of the two previous cases can be similarly generalized to:

$(\exists_D x)(\forall_A y)(Cxy)$: Some drug cures every ailment, where $I = (2^n - 1)^m\ (2^{-mn})$, i.e. about 0.91 in the present case.

One final assertion is a universal generalization of both of these cases:

$(\forall_D x)(\forall_A y)(Cxy)$: Every drug cures every ailment, where $I = 1 - 2^{-mn}$, i.e. about 0.9999 in the present case.

The inverse hierarchy occurs for the negations of these assertions. Similar hierarchies of generalization can be generated for quantifications of ternary relations, and so on. It would be an interesting exercise to determine whether logically-untrained individuals have an intuitive grasp of the relative informativeness of these assertions.

2. Probability and the strength of inductive inferences

The most frequent complaint made by the commentators was that the target article made no reference to the relation between induction and probability—indeed, the theory offered no account of probabilities (Cohen, Over & Manktelow, Newstead), it should be augmented by explanations of probabilistic reasoning (Hunt) and of probabilistic claims based on induction (Green), and it should account for the strength of inductions (Smith).

The omission struck Cohen as remarkable. He writes: "To present an account of induction that speaks only of classificatory outcomes, not of comparative or quantitative ones, is like presenting a theory of mechanics in which objects are either at motion or at rest, rather than a theory in which objects are moving at a measurable velocity relative to a chosen reference-point". Yet, the omission is justifiable. There are good philosophical precedents for discussing induction without alluding to probabilities—indeed, neither Hume's (1748) classical riddle of induction nor Goodman's (1965) new riddle depends on probabilities. Moreover, the goal of the target article was to give an account of everyday inductive thinking—the sort of inductions that are made by everyone from Aristotle to aboriginals, whether or not they know anything of the probability calculus. As far as one can tell, these inductions are part of a universal human competence: they do not depend on any overt mastery of quantities or even of natural numbers. Hence, the goals of a psychology of induction are, at least initially, quite different from those of a philosophical account. Aristotle's notion of probability amounts to the following: a probability is a thing that happens for the most part, and conclusions that state what is probable must be drawn from premises that do the same (see Rhetoric, I, 1357a). In comparison to, say, Pascal's account this notion is appallingly crude, but it corresponds to about the level of competence a psychological theory of induction should initially aspire to explain. Of course many individuals do encounter the probability calculus at

school; few master it. A simple observation shows an inadequate grasp of the fundamentals. One poses the following problem:

> There are two events, A and B, which each have a probability of a half. What is the probability that A and B both occur?

Many people respond: a quarter. In this case, the appropriate 'therapy' is to invite them first to imagine that A is a coin landing heads and B is the same coin landing tails, i.e. $p(A \& B) = 0$, and then to imagine that A is a coin landing heads and B is a coin landing with the date uppermost, i.e. $p(A \& B) = 0.5$. At this point, most people began to grasp that there is no definite answer to the question posed in the puzzle—joint probabilities depend on the dependence of one event on the other. Other phenomena show that intelligent individuals do not know how probabilities combine according to logical operations. Indeed, we are all likely to go wrong in thinking about probabilities: the calculus is a piece of technology that few people completely master.

Another reason for treating induction separately from probability is that many probabilistic inferences are deductive rather than inductive (*cf.* Aristotle's view). Here, for example, is a piece of probabilistic reasoning:

$p(A) = 0.5$
$p(B/A) = 1$ (i.e. the conditional probability of B given A is 1)
$p(B/\text{not } A) = 0$
$\therefore\ p(B) = 0.5$

The inference is a deduction. It makes explicit what is implicit in the premises, and it does not increase their semantic information. Here is a more mundane example inspired by Over & Manktelow's discussion:

> If my son smokes, he is in danger of damaging his health.
> My son smokes.

Suppose you know the first of these premises for certain, but your degree of belief in the second premise falls short of full certainty. What, Over and Manktelow ask, should you infer? How will you infer it? And what degree of confidence should you assign to the conclusion? In fact, the following inference is a valid deduction, which can be drawn with certainty:

> If my son smokes, he is indanger of damaging his health.
> My son probably smokes.
> $\therefore$ My son is probably in danger of damaging his health.

Over & Manktelow claim that mental models can be used only to represent alternative states of affairs that are treated as equally likely. In fact, as we will see, there is no reason to suppose that when you compare models you take them to be equally likely. To illustrate the point, consider another example suggested by Newstead's comment that probabilistic inference may be hard to incorporate into the model theory:

> Most archers are broad-shouldered.
> Robin is one of the archers.
> $\therefore$ Probably, Robin is broad-shouldered.

Again, this inference is a deduction, not an induction, and it is easily accommodated within the theory. The key step is to construct models that represent proportions in order to cope with quantifiers such as 'most' (see Johnson-Laird, 1983, p. 137). Next, following Aristotle, assertions of the form: *probably S*, are treated as equivalent to: *in*

most possible states of affairs, S. Thus, a model of the assertion that most archers are broad-shouldered takes the form:

```
[a]   b
[a]   b
[a]   b
[a]
```

where the set of archers is exhaustively represented. When the information from the second premise is added to this model, one possible state of affairs is:

```
r   [a]   b
    [a]   b
    [a]   b
    [a]
```

in which Robin is broad-shouldered. Another possibility is:

```
    [a]   b
    [a]   b
    [a]   b
r   [a]
```

in which Robin is *not* broad-shouldered. In most possible states of affairs, however, Robin will be broad-shouldered. Hence, one can deduce: probably Robin is broad-shouldered. This same method accounts for the inference about the smoker. Individuals who are capable of one-to-one mappings but who have no access to cardinal or ordinal numbers will still be able to make this inference. They have merely to map each possible state of affairs in which S occurs one-to-one with each possible state of affairs in which S does not occur: if there is a residue, then it corresponds to the more probable category.

Over and Manktelow suggest that model elimination as the basis of induction is too narrow. They write: "Even if premises are held with certainty, such reasoning should not utterly eliminate alternative models. The alternatives should only be believed or held probable to a different (lesser) degree". Newstead similarly remarks that it is difficult to see how models can represent information other than in an all-or-none fashion. There are a number of misconceptions that need to be separated here. Induction does not necessarily lead to the elimination of a model—it is a process of adding information to models, which sometimes leads to the elimination of a model. Similarly, as the previous example shows, the process of adding information to a model may lead to alternative possibilities and thus to a probabilistic conclusion.

Of course, certain *inductions* are probabilistic too. So what underlies these inferences and renders them strong or weak? Johnson-Laird (in press) outlines a model-based theory of probabilistic induction and shows how the strength of an inference is accounted for within this theory. The next section will outline these ideas, and subsequent sections will use them to answer commentators' specific questions.

Models, strength, and inductive probability

Although, by definition, inductive arguments are all logically invalid, they differ in their strength—some are highly convincing, others are scarcely credible, at least in relation to

a given background of knowledge. The manifest differences in strength are an important clue about the psychology of inductive inference. If we could understand what determines the strength of an induction, we would have made progress towards a psychological theory of what the mind computes in inference, i.e. a theory at the 'computational' level. We need to distinguish between the strength of an argument—the degree to which its premises, if true, support the conclusion, and the degree to which the conclusion is likely to be true—a notion for which we need a theoretically neutral term: 'probability' suggests an immediate tie to the probability calculus, and so we will refer instead to the 'credibility' of propositions. An argument can be strong but its conclusion incredible because the argument is based on incredible premises, e.g. a valid argument based on false premises can lead to a false conclusion. Hence, we will distinguish between the credibility of the premises and the strength of the argument, and we will propose that in principle the credibility of a conclusion should depend on both the credibility of the premises and the strength of the argument. As we shall see, individuals are liable to neglect the second of these components.

Readers will have noted that we talk of both inductive inference and inductive argument. They are one and the same thing, but the two terms bring out the point that the informal arguments of everyday life, which occur in conversation, newspaper editorials, and scientific papers, often hinge on inductive inferences. The strength of such arguments depends on the relation between the premises and the conclusion. But the nature of this relation is deeply puzzling. The puzzle is so great that many theorists have abandoned logic altogether in favor of their own idiosyncratic methods of assessment (see e.g. Toulmin, 1958; and many other discussions of informal argument). Other accounts rely on human beings to determine which assertions support which other assertions (e.g. Thagard, 1989). And still others use rules to which they assign a numerical parameter corresponding to certainty (see Part 3).

Osherson, Smith and Shafir (1986) in a ground-breaking analysis explored a variety of accounts of inductive strength that boil down to three main hypotheses: (1) an inference is strong if, given an implicit assumption, schema, or causal scenario, it is logically valid, i.e. the inference is an enthymeme (cf. Aristotle); (2) an inference is strong if it corresponds to a deduction in reverse, such as an argument from specific facts to a generalization of them (cf. Hempel); and (3) an inference is strong if the predicates (or arguments) in premises and conclusion are similar (cf. Tversky and Kahneman: as Smith rightly observes the issue of strength must relate to these authors' research into heuristics). Each hypothesis has its strengths and weaknesses, but their strong points can be captured in the following proposals, which we will develop in two stages. First, we will specify an abstract characterization of *what* in principle has to be computed in order to determine the strength of an inference (i.e. a theory at the 'computational' level); and, second, we will specify *how* in practice the mind assesses the strength of an argument (i.e. a theory at the 'algorithmic' level).

The relation between premises and conclusion is a semantic one, and it can be characterized abstractly by adopting the semantic approach to logic (see e.g. Barwise & Etchemendy, 1989). A set of premises, including implicit premises provided by general knowledge, lend *strength* to a conclusion according to two principles, which depend on considering all infinitely many possible states of affairs consistent with the premises:

(1) The conclusion is true in at least one of the possible states of affairs in which the premises are true, i.e. the conclusion is at least consistent with the premises. If there is

no such state of affairs, then the conclusion is inconsistent with the premises: the inference has no strength whatsoever, and indeed there is valid argument in favor of the negation of the conclusion.

(2) Possible states of affairs in which the premises are true but the conclusion false (i.e. counterexamples) weaken the argument. If there are no counterexamples, then the argument is maximally strong—the conclusion follows validly from the premises. If there are counterexamples, then the strength of the argument is equal to the proportion of states of affairs in which the premises and conclusion are true.

This account has several advantages.

First, it does not throw the "logical" baby out with the bath water. What underlies deduction is the semantic principle of validity: an argument is valid if its conclusion is true in any state of affairs in which its premises are true. The present account does not abandon this principle merely because inductive inferences are invalid and so cannot be captured by valid formal rules of inference. By definition, an induction increases semantic information and so its conclusion must be false in possible cases in which its premises are true. Hence, inductions *are* reverse deductions, but they are the reverse of deductions that throw semantic information away.

Second, if each possible state of affairs is assumed to be equi-possible, then the addition of a measure of the cardinality of the relevant sets of states of affairs provides an extensional foundation for probabilities, i.e. the strength of an inductive argument is equivalent to the probability of the conclusion given the premises. But the two abstract principles are not equivalent to the probability calculus: the human inferential system can attempt to estimate the relevant proportion without necessarily using the probability calculus. Likewise, the assumption that possible states of affairs provide a foundation for probabilities has no strong implications for the correct interpretation of the probability calculus, which is a matter for self-conscious philosophical reflection. The assumption is compatible with interpretations in terms of limiting frequencies of events, in terms of equi-possibilities based on physical considerations, and in terms of subjective degrees of belief (cf. e.g. Hintikka's, 1962, analysis of beliefs in terms of possibility). Hence, an argument (or a probability) may concern either a set of events or a unique event. Over and Manktelow claim: 'subjective probability is explained in terms of degree of belief, with the greatest degree being certainty'. But, individuals who are innumerate may not assign a numerical degree of certainty to their conclusion, and even numerate individuals may not have a tacit mental number representing their degree of belief. Individuals' beliefs do differ in subjective strength, but it does not follow that such differences call for a numerical system of subjective probabilities. An alternative conception of 'degrees of belief' might be based on analogue representations (cf. Hintzman, Nozawa & Irmscher, 1982), or on a system that permitted only partial rankings of strengths, such as one that recorded the relative ease of constructing different classes of models.

Third, the account is entirely compatible with semantic information. As we have seen, the semantic information conveyed by a proposition, A equals $1 - p(A)$, where 'p(A)' denotes the probability of A. If A is a complex proposition containing conjunctions, disjunctions, etc., its probability can be computed in the usual way according to the probability calculus. Induction with probabilities remains a matter of increasing semantic information (*pace* Over & Manktelow).

Fourth, the account shows how the model theory extends from deduction to everyday reasoning and argumentation—a matter raised by both Garnham and Green. One feature of such informal argumentation is that it typically introduces both a case for

a conclusion and a case against it—a procedure that is so unlike a logical proof that many theorists have supposed that logic is useless in the analysis of everyday reasoning (see e.g. Toulmin, 1958, for a highly influential account). The *strength* of an argument, however, can be straightforwardly analyzed in the terms described above: informal argumentation is typically a species of induction, which may veer at one end into deduction and at the other end into a creative process in which one or more premises are abandoned.

The disadvantage of the account is that it is obviously impossible for the mind to consider all the infinitely many states of affairs consistent with a set of premises. So how does this account translate into a psychological mechanism for assessing the strength of an argument? It is this problem that the theory of mental models is designed to solve (see Johnson-Laird, in press).

The essence of the theory is that inference depends on constructing a model based on comprehension and general knowledge, formulating a conclusion that holds in the model if none is provided, and searching for alternative models of the premises that render the putative conclusion false. A model has a structure that corresponds to the structure of states of affairs, but represents a *class* of states of affairs—a class that may have potentially an infinite number of members (Barwise, 1993). The *strength* of an influence depends, as we have seen, on the relative proportions of two sorts of states of affairs: those in which the conclusion is true and those in which it is false. Reasoners can estimate these proportions by constructing models of the premises and attending to the proportions with which the two sorts of models come to mind, and perhaps to the relative ease of constructing them. For example, given that someone fell (without a parachute) from an airplane flying at a height of 2000 feet, then they probably died. The inference is strong, but not irrefutable. One may have heard of cases to the contrary, or can imagine them—the individual falls into a large haystack, or a deep snow drift. But, in constructing models (of classes of possibilities), those in which the individual is killed will occur much more often than those in which he survives (cf. the inference above about whether Robin, the archer, is broad-shouldered). This account is entirely compatible with the idea of estimating likelihoods in terms of scenarios, which was proposed by Tversky and Kahneman (1973, p. 229), and it forms a bridge, as Smith requested, between the model theory and the heuristic approach to judgements of probability. Estimates of the relative frequencies of the two sorts of models—those in which the conclusion is true and those in which it is false—will be rudimentary, but they should be less biased by the sorts of factors to which Smith alludes than estimates of probability based on the typicality of a single exemplar (Kahneman, personal communication).

The strength of an argument depends on the relation between the premises and the conclusion, and, in particular, on the proportion of possibilities compatible with the premises in which the conclusion is true. This relation is *not* in general a formal or syntactic one, but a semantic one. It takes work to establish the proper relation, and the theory makes a number of predictions about making and assessing inductive inferences. First, arguments—especially in daily life—do not wear their logical status on their sleeves, and so individuals will tend to approach deductive and inductive arguments alike. They will tend to confuse an inductive conclusion, i.e. one that could be true given the premises, with a deductive conclusion, i.e. one that must be true given the premises. Second, envisioning models, which each correspond to a class of possibilities, is a crude method, and, because of the limited processing-capacity of working memory, many models are likely never to be envisaged at all. The process will be affected by the constraints that were mentioned in the target article: specificity, parsimony, and

availability of relevant knowledge. Third, individuals are likely to be inferential satisficers, i.e. if they reach a *credible* conclusion, or succeed in constructing a model in which such a conclusion is true, they are likely to accept it, and to overlook models that are counterexamples. Conversely, if they reach an *incredible* conclusion, they are likely to search harder for a model of the premises in which it is false. Fourth, the propensity to satisfice will in turn lead them to be overconfident in their conclusions, especially in the case of arguments that do have alternative models in which the conclusion is false. Fifth, the process of assessment—the construction of models—relies on heuristics. These heuristics, which have been extensively explored by Tversky and Kahneman, can be traced back to Hume's seminal analysis of the connection between ideas: 'there appear to be only three principles of connexion between ideas, namely, *Resemblance, Contiguity* in time or place, and *Cause* or *Effect*' (Hume, 1748, Sec. III). Hence, as Smith points out, one cue is the semantic similarity between the premises and the conclusion, and another cue is the causal cohesiveness between them. They are also likely to rely more on the credibility of premises (and conclusion) than on the strength of the argument, i.e. the relation between the premises and conclusion. Sixth, individuals are likely to focus on what is explicit in their initial models and thus to be susceptible to various 'focusing effects' (see Legrenzi, Girotto & Johnson-Laird, 1993). Finally, the construction of an explanatory model provides an argument of greater strength than a mere inductive generalization, because an explanation demonstrates the impossibility of the cause not leading to the consequence. In the next section, we will use the present theory to answer the commentators' specific questions about predictions.

Some queries about the model theory's predictions

Cohen writes: 'we want to know whether inductive support for the generalization that all As are B increases with the number of As that are known to be Bs and, if so, whether or not the rate of increase is constant, whether or not the rate of increase is affected by relevant differences between the instances, and whether or not this kind of relevance is itself a matter of degree'. Newstead also asks how many instances are needed for a generalization. According to the model theory, no number of observations of As that are Bs suffices for the generalizations that all As are Bs if one can readily envisage As that are not Bs. Other things being equal, these proportions should be reflected in individuals' judgements about the relation between As and Bs. Of course, as the target article pointed out, reasoners bring so many assumptions to inductive reasoning that a single instance may suffice for a strong conclusion. One experience of a wheel clamp on your car may lead to the generalization that you are likely to be clamped again if you park on double yellow lines in London.

The model theory answers another of Cohen's questions: how do logical operations on a hypothesis, such as conjunction, disjunction, contraposition, affect the value of inductive support? The classic case is Tversky and Kahneman's (e.g. 1983) demonstration of the 'conjunction fallacy', i.e. a violation of the elementary principle that $p(A \& B) \leqslant p(B)$. For example, a woman who is described as 31 years old, liberal, and outspoken, is judged more likely to be a feminist bankteller than a bankteller. Smith argues cogently:

> The description of the woman might lead to a model of her that includes the additional information that she has feminist beliefs; because this model pro-

vides a closer match to feminist bankteller than to bankteller, the former alternative is favored as an inference.

There is whole battery of 'focusing' effects that can be explained by the model theory. Consider, for instance, a task used by Beyth-Marom and Fischhoff (1983). They presented subjects with the following scenario: "You have met Mr Maxwell at a party to which only university professors and business executives were invited. The only thing you know about Mr Maxwell is that he is a member of the Bear's Club". At this point one group of subjects had to assess the probability of Maxwell being a university professor, while a second group had to decide whether it was more probable that he was a university professor than a business executive. They also rated the relevance of several questions which they could ask in order to make their judgements. Most subjects in both groups rated as relevant the proportion of professors who were members of the club. In contrast, significantly fewer subjects in the first group than in the second group rated as relevant the proportion of *executives* who were members of the club. The difference demonstrates an inability to grasp which data are diagnostic. It is akin to the "pseudo-diagnosticity" bias postulated by Doherty, Mynatt, Tweeney and Schiavo (1979). Their subjects had to decide whether a pot came from island A or to island B. Once they formed an hypothesis about the origin of the pot—say, island A—they focused on information concerning that hypothesis (how many features of the pot were present in the pots of island A) and ignored the alternative hypothesis (how many features of the pot were present in the pots of island B). The phenomena seem to be a consequence of the way in which subjects build mental models. When the judgement is between two alternatives, as for the second group in Beyth-Marom and Fischhoff's experiment, reasoners build two alternative models (professor and executive) and hence ask for information about them. When subjects are focused on a specific target (professors, pot A), they construct only a single model and so fail to consider relevant information concerning alternative hypotheses.

The inevitable tendency to focus on what is explicit in models also accounts for the 'positive test' strategy, i.e. the preference for testing positive instances of a hypothesis (Klayman & Ha, 1987). It may also contribute to 'framing' effects in decision making (see e.g. Tversky & Kahneman, 1981): equivalent descriptions of the same decision leading to different initial models can elicit different patterns of choice. Comparable effects should also occur in estimates of the probabilities of various events. Consider, for example, the following two descriptions:

(A) If there is a short circuit then there's an increase in power.
(B) There's a short circuit only if there's an increase in power.

In which case is there more likely to be both a short circuit and an increase in power? Likewise, in which of these two cases is it more likely that one or other of the two events does not occur?

(C) If there isn't a cut in spending, then there's an increase in consumption.
(D) If there is a cut in spending, then there isn't an increase in consumption.

Even psychologists are apt to be confused about these matters. In fact, the first pair of descriptions are logically equivalent, but the model theory predicts that first will be

judged to lead to the two events more often than the second. The initial models of the conditional A are as follows:

```
[s]  p
 ...
```

where s denotes a short circuit and p denotes the increase in power, and so the models make explicit only the joint occurrence of the two. The initial models for the "only if" description B make explicit the negative contingency that without the increase in power there is no short circuit (Johnson-Laird & Byrne, 1991).

```
  [s]     p
¬ s  [¬ p]
 ...
```

Yet, the two sets of models are exactly the same when the implicit models are fleshed out. The theory predicts that subjects should be indifferent between C and D, because their initial models contain one positive and one negative element. The description in C yields the initial models:

```
[¬s] c
 ...
```

where s denotes a cut in spending and c denotes an increase in consumption. The description in D yields the initial models:

```
[s]  ¬c
 ...
```

In fact, one or other of the two events is more likely *not* to occur in case D than in case C, but the difference emerges only when the models are fleshed out explicitly. C yields:

```
¬ s   c
  s   c
  s  ¬ c
```

in which one of the two events always occurs. D yields:

```
  s   ¬ c
¬ s     c
¬ s   ¬ c
```

in which there is one contingency in which neither of the two events occur. In short, subjects should find it difficult to consider all the alternative models corresponding the different descriptions. They will make their judgements on the basis only of the initial models, and so they will erroneously prefer option A to option B where there is in reality no difference between them, and they will fail to prefer option D to option C where there is in reality a genuine difference between them.

For the same reason, logically-untrained individuals should not readily grasp that a conditional and its contrapositive express propositions with the same truth conditions. In seeking or assessing evidence for the claims, they will accordingly act in rather different ways given the model theory's assumption that the process is governed by what is explicit in models. They will not be tempted to treat 'all ravens are black' as equivalent to 'all non-black entities are non-ravens'.

Cohen raises one final question: how can a theory of induction based on models deal with Goodman's (1965) new riddle of induction? The essence of the riddle can be paraphrased as follows (Johnson-Laird, 1988, p. 235): You observe a series of cases of

smallpox and notice that each patient had a prior contact with someone suffering from the disease. You draw the inductive conclusion:

> If anyone is in contact with a case of smallpox they are likely to catch the disease.

(Of course, you do not know whether there are individuals who had the contact but failed to develop the disease.) But, as Goodman's riddle makes clear, your evidence also supports the conclusion:

> If anyone is in contact with a case of smallpox, then until the year 2000 they are likely to catch the disease, and thereafter they are likely to catch measles.

This inference is silly, but why? You may say: because we know that diseases no more change their spots than leopards do. But how do you know that? You may say: because all your observations support this claim. But, of course, all your observations are equally consistent with the claim that both leopards and diseases will change their spots in the year 2000. It seems, as Goodman concludes, we have no way to distinguish between 'law-like' generalizations and 'accidental' ones, and certainly no way in which to do so in terms of logical form. Goodman goes on to develop a solution in terms of a knowledge of past regularities (and of how they are described linguistically) and in particular the 'projectability' of predicates. The psychological problem, however, is perhaps simpler: people generalize on the basis of their previous knowledge of entities and their properties (regardless of whether they are justified in doing so by Goodman's principles). Indeed, as the target article points out, the price of induction is imperfection—the fads of pseudo-science, superstitions, and 'magical' thinking, which with hindsight are almost as ridiculous as those generalizations in which entities change their properties in the year 2000.

Individuals are often over-confident in their inductive judgments, and Over and Manktelow contrast the model theory unfavorably with the theory of 'probabilistic mental models' propounded by Gigerenzer, Hoffrage and Kleinbölting (1991), which they take to be of far more use in accounting for over-confidence. In fact, probabilistic mental models are intended to account for inductive answers to questions, i.e. choices between alternatives, but they are certainly not designed to explain either induction or probabilistic inference in general. Moreover, they presuppose that individuals build up a knowledge of probabilistic cues and their validities (in the form of conditional probabilities), and that they choose answers and judge their confidence using the single cue with the strongest validity and without any aggregation of multiple cues. The principle prediction of the theory is that confidence derives from the validity of the strongest cue, and the authors report corroboratory evidence from their experiments on the phenomenon of over-confidence, i.e. rated confidence tends to be higher than the actual percentage of correct answers. As Griffin and Tversky (1992) point out, however, over-confidence is greater with harder questions and this factor provides an alternative account of Gigerenzer *et al.*'s results.

What does the present theory have to say about over-confidence in induction? The propensity to satisfice (see the previous section) should lead subjects to overlook models unwittingly in the case of multiple-model problems, and so they should tend to be more confident than justified in the case of these problems. With easier one-model problems, the error and its correlated over-confidence cannot occur. Once again, this account is largely in agreement with the heuristic approach. Griffin and Tversky (1992) distinguish between the size of an effect (e.g. the difference in means) and its significance (e.g. as

dependent on sample size). They argue that individuals fail to combine the two according to statistical principles, but rather concentrate on size and then fail to adjust it adequately for significance. According to the model theory, individuals build an initial model that makes explicit the case for a conclusion, and then fail to adjust their estimates of its likelihood by taking into account alternative possibilities. In the unpublished study by Johnson-Laird and Anderson, which was mentioned in the target article, subjects were asked to draw initial conclusions from such premises as:

> The old man was bitten by a poisonous snake. There was no known antidote available.

They tend initially to infer that the old man died. Their confidence in this conclusion was moderately high. They were then asked whether there were any other possibilities and they usually succeeded in thinking of two or three. When they could go no further, they were asked to rate again their initial conclusions, and showed a reliable decline in confidence. Hence, by their own light, they were initially overconfident. The theory certainly predicts that over-confidence should be a function of difficulty, because easy problems depend on only a single model. But should subjects be underconfident in such cases, as is sometimes observed? One factor that may be responsible for the effect in repeated-measure designs is the subjects' uncertainty about whether or not there might be other models in a one-model case.

In summary, the model theory is neither neutral on the questions the commentators raise nor is it empirically untestable. At the computational level, it is compatible with a normative account based on the probability calculus (see Carnap, 1950; Hesse, 1974), and it allows for the development of the calculus as a formal exercise of mathematical thinking. The theory at the algorithmic level, however, does not imply that statistically-naive individuals compute numerical probabilities explicitly, or that they combine them according to the rules of probability calculus.

Constraints on models

The target article postulated that induction is a process of adding information to models, and that the process is constrained by a number of factors: existing data, specificity (keeping the model as specific as possible), parsimony, the availability of background knowledge. Hunt asks: how are candidate assertions to be added to models generated in the first place? Strictly speaking, it is information rather than assertions per se that is added to models, and this information derives from the constraints: induction is a constraint-satisfaction process. Hunt goes on to suggest that schema application and Bayesian reasoning may also have roles to play. The schemas that one has, say, for detective stories (as in Hunt's example) or any other domain, can be subsumed under the general principle of background knowledge. Hunt also suggests that people examine the most probable causes first. But how do they recover or assess the most probable cause? Once again, the process is likely to hinge on the availability of knowledge. Mental models (unlike computer implementations of the theory) are *semantic* representations, and the theory recognizes the importance of knowledge in reasoning. Yet, reasoning is more than just knowledge: the present author is less tempted than Hunt to hand cognitive psychology over to anthropologists!

Bara accepts the general thesis of the model theory, but expresses strong reservations about the availability heuristic. Which notions are the available ones in a particular context? Surely available knowledge has also to be relevant and pertinent? What is the

mechanism that activates a specific piece of general knowledge? In fact, the model theory postulates that the mechanism relies on a series of model-based strategies for making knowledge progressively available. This hypothesis provides a vital, but hitherto missing, part of the theory of informal inference. Reasoners begin by trying to form a model of the current situation, and the retrieval of relevant knowledge is easier if they can form a single model containing all the relevant entities. They do not rely on a linguistic description of the situation, which could use only 'key' words that occur in it as a basis for retrieval. Such a system would probably be unworkable and psychologically implausible, triggering either too much knowledge or not enough. Once reasoners have formed an initial model, knowledge becomes available to them in a systematic way. They manipulate the spatial or physical aspects of the situation, i.e. they manipulate the model directly by procedures corresponding to such changes. Next, they make more abstract conceptual manipulations, e.g., they consider the properties of superordinate concepts of entities in the model. Finally, they make still more abstract inferences based on introducing relations retrieved from models of analogous situations (cf. Gentner, 1983). Consider the following illustration:

> Arthur's wallet was stolen from him in the restaurant. The person charged with the offense was outside the restaurant at the time of the robbery. What follows?

Reasoners are likely to build an initial model of Arthur inside the restaurant when his wallet is stolen and the suspect outside the restaurant at that time. They will infer that the suspect is innocent. They may then be able to envisage the following sort of sequence of ideas:

(1) Physical and spatial manipulations:

> The suspect leant through the window to steal the wallet.
> The suspect stole the wallet as Arthur was entering the restaurant, or the thief ran in and out of the restaurant very quickly [ideas that, in fact, are contrary to the premises—as informal inferences quite often tend to be].
> The suspect used a device on a long pole to reach in through the window to steal the wallet.

(2) Conceptual manipulations:

> The suspect had an accomplice—a waiter, perhaps—who carried out the crime [theft is a crime, and many crimes are committed by accomplices].

(3) Analogical thinking:

> The suspect used a radio-controlled robot to sneak up behind Arthur to take the wallet [by analogy with the use of robots in other "hazardous" tasks].

In short, the theory predicts that reasoners begin by focusing on the initial explicit properties of a model, and then they attempt to move away from them, first by conceptual operations, and then by introducing analogies from other domains. It is important to emphasize that the order of the three sorts of operations is not inflexible, and that particular problems may elicit a different order of operations. Nevertheless, there should be a general trend in moving away from the explicit model to more remote possibilities.

3. Are there rules for induction?

This question is raised by several of the commentators, notably by Girotto, Hunt and Smith. Hunt describes his AI program, the reactive library, that draws inductions such as:

Strychnine facilitates learning in rats.
Strychnine facilitates learning in mice.
∴ Strychnine facilitates learning in rodents.

Such inferences depend on a hierarchy of class-inclusion relations, and are based on the rule of drawing the least general conclusion that subsumes the facts. Hunt argues that human induction depends on more than just such rules: it depends on individuals' theories of the world. The point is well-taken. As I wrote in the target article: "induction is a search for a model that is consistent with observation and background knowledge". Indeed, I went on to argue:

> The most important constraint on induction I have left until last for reasons that will become clear. It is the use of existing knowledge. A rich theory of the domain will cut down the number of possible inductions; it may also allow an individual to generalize on the strength of only a single instance.

Hunt in paying credit where it is due attributes the idea to Murphy and Medin (1985), but they in turn pay credit to Miller and Johnson-Laird (1976), who argued that concepts themselves embody proto-theories. In any case, Hunt is right to emphasize that knowledge, especially in the organized form of strong theories or models, exerts an essential constraint on inductive reasoning.

Collins and Michalski (1989) have developed a more sophisticated version of rule-based induction: they argue, like Holland, Holyoak, Nisbett and Thagard (1986) that individuals construct models on the basis of rules of inference, and that these rules have numerical parameters governing such matters as certainty. They have not tried to formalize all patterns of plausible inference, but rather some patterns of inference that make up a core system of deductions, analogies, and inductions. The main novelties of their approach are the use of a formal language that allows relations between variables to be stated (e.g. the latitude of a place is inversely related to its average temperature), and the use of parameters that express such matters as the certainty of a statement, the typicality of an instance of a member of a set, and so on. Their system includes a set of eight rules for transforming one statement into another on the basis of class-inclusion information of the sort to be found in a semantic network. Hence, the inference:

The flowers of England include daffodils and roses.
∴ The flowers of Europe include daffodils and roses.

is made by such a transform given that the semantic network represents the relevant facts that Europe generalizes England in the context of climate, and that climate determines flora. The formal specification of this transform is as follows:

d(a) = r	[flowers (England) = {daffodils, roses ...}]
a′ GEN a in CX (a′, D(a′))	[Europe generalizes England in the context of climate]
D(a′) <------> d(a′)	[Climate determines what flowers grow]
∴ d(a′) = r	[flowers (Europe) = {daffodils, roses ...}]

People are supposed to make such inferences provided they have no contrary information, but the certainty of a conclusion can depend on the values of seven parameters. Every statement transform depends on a mutual dependency (as shown by the double-headed arrow), and the greater the conditional likelihood between the variables, the greater the certainty in the inference. The other parameters include similarity between concepts, the typicality of the argument as a member of its superset, e.g. England of Europe, the frequency of a reference set within the argument set, and the multiplicity of the referent and of the argument.

Collins and Quillian also propose a further nine rules of inference for making inferences from propositions that mutually imply one another and from mutual dependencies between variables. Here is an example based on one such rule concerning a mutual implication (in the first premise);

> If a place has a warm climate, a heavy rainfall, and a flat terrain, then it can grow rice.
> Florida is a place.
> Florida has a warm climate.
> Florida is flat.
> Uncertain whether Florida has a heavy rainfall.
> Therefore, if Florida has a heavy rainfall, it can grow rice.

Apart from the parameters concerning certainty, the same inference can be made as a valid deduction in the propositional calculus without the need for a special rule of inference tailor-made for it. Hence, a general theory of reasoning with quantifiers and connectives obviates the need for this and other deductive transforms in Collins and Quillian's account, provided that it offers an account of the strength of an argument.

Collins and Michalski (1989, p. 40) state that it is difficult to use standard psychological techniques to test their theory. The theory is intended to account only for people's answers to questions. It does not make any predictions about the differences in difficulty between various sorts of inference, and, as they point out (p. 7), it does not address the issue of whether people make systematic errors. Hence, their main proposed test consists in trying to match protocols of arguments against the proposed forms of rules. Pennington and Hastie (1993) report success in matching these patterns to informal inferences of subjects playing the part of trial jurors. But, as Collins and Michalski mention, one danger is that subjects' protocols are merely rationalizations for answers arrived at by other means. Another difficulty is that there are no definitive criteria for what counts as a match between a protocol and a rule of inference. Anyone who has had experience in translating everyday language into a logical notation knows that it is all too easy to analyze everyday expressions into many different formal patterns. In sum, AI rule systems for induction are under development, but so far they have not received any very striking empirical corroboration.

In contrast, as Smith says, another sort of rule theory has much more empirical support. This theory appeals to the idea that individuals have a tacit knowledge of such rules as the 'law of large numbers'. Girotto argues that the empirical evidence appears strongly to support the use of such rules, and thus constitutes a challenge to the model theory. Individuals apply the rules to novel materials, mention them in justifying their responses, benefit from training with them, and sometimes overextend their use of them. 'All these pieces of evidence,' Girotto writes, 'seem to support rule-theories of evidence' (see also Smith, Langston, and Nisbett, 1992). It is instructive, however, to compare

these rules with those to be found in AI systems. Here is an example of an inductive AI rule (see Fig. 1 in the target article):

If p & q then s
∴ If p then s

This rule is formal and can be applied to the representation of the abstract logical form of premises. The law of large numbers can be paraphrased as follows:

> The larger the sample from a population the smaller the tendency for the sample mean to diverge from the population mean.

Although Aristotle may not have grasped at once such notions as sample mean and population, he would probably have been more surprised by a married couple having 10 children who were all girls than by a couple with three children who were all girls. He would thus have had a tacit grasp of the law, which he could make use of in certain circumstances. The law of large numbers, however, is *not* a formal rule of inference. It has a rich semantic content that goes well beyond the language of logical constants, and it is doubtful whether it is applied to the logical form of premises. On the contrary, the law is only likely to be applied when one has grasped the content of a problem, i.e. constructed a model of it. Yet, the law is a general principle that can be applied to many different situations.

There are many other general principles of general knowledge, e.g. to multiply by 10 add 0 to the decimal numeral, to get out of certain mazes keep turning left, drive on the left in Japan. They differ in generality and in validity, but they certainly exist. The fact that individuals can be taught such rules and that they sometimes err in overextending them tells us nothing about their format. They may take the form of schemas or content-specific rules of inference, but they could be represented declaratively. Likewise, how they enter into the process of thinking—the details of the computations themselves—is also not known. There is, however, no reason to oppose them to mental models. They seem likely to work together in tandem, just as conceptual knowledge underlies the construction of models.

4. Conclusions

Girotto argues generously that the model theory presents the most complete account of human deduction, and Garnham points out that it was never intended to be just a theory of this domain. It may provide real hope, he argues, for a theoretical framework that will impose order on the study of thinking and reasoning. Similarly, Green emphasizes that different sorts of reasoning should be able to operate in the same mental workspace: induction may, as Aristotle argued, provide the premises for a deduction. Problem solving and everyday reasoning may bring together all the different varieties of thought. What the target article introduced was a taxonomy of the different steps in thought with a propositional content. Mental models may be states in a problem-space—the conception that Garnham advocates, and that both Herb Simon (personal communication) and the late Alan Newell (1990) have defended. Paradoxically, given the interdependence of process and representation, it has proved to be much harder to pin down cognitive processes than cognitive representations. Production systems and even certain parallel network systems have universal Turing machine power, and hence, as Garnham reminds us, their explanatory power often depends on the specific accounts

framed within them. The evidence from deductive reasoning strongly supports a system based on mental models. What this reply to the commentators has tried to show is that mental models are an equally feasible framework for induction.

References

BARWISE, J. (1993) Everyday reasoning and logical inference. (Commentary on Johnson-Laird & Byrne: Deduction) *Behavioral and Brain Sciences*, 16, pp. 337–338.

BARWISE, J. & ETCHEMENDY, J. (1989) Model-theoretic semantics. In POSNER, M.I. (Ed.) *Foundations of Cognitive Science* (Cambridge, MA, MIT Press).

BEYTH-MAROM, R. & FISCHHOFF, B. (1983) Diagnosticity and pseudo-diagnosticity. *Journal of Personality and Social Psychology*, 45, pp. 1185–1195.

CAREY, S. (1985) *Conceptual Change in Childhood* (Cambridge, MA, MIT Press).

CARNAP, R. (1950) *Logical foundations of Probability* (Chicago, Chicago University Press).

COLLINS, A.M. & MICHALSKI, R. (1989) The logic of plausible reasoning: A core theory. *Cognitive Science*, 13, pp. 1–49.

DOHERTY, M.E, MYNATT, C.R., TWENEY, R.D. & SCHIAVO, M.D. (1979) Pseudodiagnosticity. *Acta Psychologica*, 43, pp. 111–121.

GENTNER, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, pp. 155–170.

GIGERENZER, G., HOFFRAGE, U. & KLEINBÖLTING, H. (1991) Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, pp. 506–528.

GOODMAN, N. (1965) *Fact, Fiction, and Forecast*, Second edition (Indianapolis, New York, Bobbs-Merrill).

GRIFFIN, D. & TVERSKY, A. (1992) The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24, pp. 411–435.

HESSE, M. (1974) *The Structure of Scientific Inference* (London, Macmillan).

HINTIKKA, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca, Cornell University Press).

HINTZMAN, D.L., NOZAWA, G. & IRMSCHER, M. (1982) Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior*, 21, pp. 127–141.

HOLLAND, J.H., HOLYOAK, K.J., NISBETT, R.E. & THAGARD, P. (1986) *Induction: Processes of Inference, Learning and Discovery* (Cambridge, MA, MIT Press).

HUME, D. (1748/1988) *An Enquiry Concerning Human Understanding* (La Salle, Illinois, Open Court).

JOHNSON-LAIRD, P.N. (1983) *Mental Models* (Cambridge, MA, Harvard University Press).

JOHNSON-LAIRD, P.N. (1988) *The Computer and the Mind* (London, Fontana) (Second edition, 1993).

JOHNSON-LAIRD, P.N. (1993) *Human and Machine Thinking* (Hillsdale, NJ, Lawrence Erlbaum Associates).

JOHNSON-LAIRD, P.N. (in press) Mental models and probabilistic thinking. *Cognition*.

JOHNSON-LAIRD, P.N. & BYRNE, R.M.J. (1991) *Deduction* (Hillsdale, NJ, Lawrence Erlbaum Associates).

KLAYMAN J. & HA, Y.-W. (1987) Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94, pp. 211–228.

LEGRENZI, P., GIROTTO, V. & JOHNSON-LAIRD, P.N. (1993) Focussing in reasoning and decision making. *Cognition*, 49, pp. 37–66.

MILLER, G.A. & JOHNSON-LAIRD, P.N. (1976) *Language and Perception* (Cambridge, MA, Harvard University Press).

MURPHY, G.L. & MEDIN, D.L. (1985) The role of theories in conceptual coherence. *Psychological Review*, 92, pp. 289–316.

NEWELL, A. (1990) *Unified Theories of Cognition* (Cambridge, MA, Harvard University Press).

NISBETT, R.E. (Ed.) (1993) *Rules for Reasoning* (Hillsdale, NJ, Lawrence Erlbaum Associates).

OSHERSON, D.N., SMITH, E.E. & SHAFIR, E. (1986) Some origins of belief. *Cognition*, 24, pp. 197–224.

PEIRCE, C.S. (1958) *Selected Writings: Values in a Universe of Chance* (New York, Doubleday).

PENNINGTON, N. & HASTIE, R. (1993) Reasoning in explanation-based decision making. *Cognition*, 49, pp. 123–163.

SMITH, E.E., LANGSTON, C. & NISBETT, R.E. (1992) The case for rules in reasoning. *Cognitive Science*, 16, pp. 1–40.

THAGARD, P. (1989) Explanatory coherence. *Behavioral and Brain Sciences*, 12, pp. 435–502.

TOULMIN, S.E. (1958) *The Uses of Argument* (Cambridge, Cambridge University Press).

TVERSKY, A. & KAHNEMAN, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, pp. 207–232.

TVERSKY, A. & KAHNEMAN, D. (1981) The framing of decisions and the psychology of choice. *Science*, 211, 453–458.

TVERSKY, A. & KAHNEMAN, D. (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, pp. 293–315.