Psychological Review
2004, Vol. 111, No. 3, 640–661

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.3.640

Reasoning From Inconsistency to Consistency

P. N. Johnson-Laird
Princeton University

Vittorio Girotto
Centre National de la Recherche Scientifique and
Venice Architecture University

Paolo Legrenzi
Venice Architecture University

This article presents a theory of how individuals reason from inconsistency to consistency. The theory is based on 3 main principles. First, individuals try to construct a single mental model of a possibility that satisfies a current set of propositions, and if the task is impossible, they infer that the set is inconsistent. Second, when an inconsistency arises from an incontrovertible fact, they retract any singularly dubious proposition or any proposition that is inconsistent with the fact; otherwise, they retract whichever proposition mismatches the fact. A mismatch can arise from a proposition that has only mental models that conflict with the fact or fail to represent it. Third, individuals use their causal knowledge—in the form of models of possibilities—to create explanations of what led to the inconsistency. A computer program implements the theory, and experimental results support each of its principles.

Reasoning is seldom a clear-cut matter of the deduction of conclusions that follow from premises. You often draw conclusions that you later withdraw in the light of new information. Suppose, for instance, that you believe the following propositions:

If Paolo has gone to get the car, then he will be back in five minutes.

and

Paolo has gone to get the car.

You think to yourself, So, he'll be back in five minutes. The inference is valid; that is, its conclusion must be true given that its premises are true. Five minutes go by, and then another 10, with no sign of Paolo. Something has to "give." You have detected an inconsistency between a valid consequence of your beliefs and a fact. You have at the very least to retract your conclusion that Paolo will be back in five minutes. But, to hold beliefs with consequences inconsistent with the facts is a hallmark of irrationality. Hence, you will probably try to reason your way to consistency. You may change your mind about whether Paolo went to get the car, or about your conditional assumption that he would be back in five minutes. This example is typical of daily life, and similar conflicts occur in science. You believe, say, that heat is a substance and that substances have weight, but then you observe that heating an object has no effect on its weight. Hence, you try to account for the inconsistency between your observations and your hypothesis. Similarly, a major search in contemporary physics is for a theory that reconciles the inconsistency between relativity theory and quantum theory (Greene, 2000). The nature of your reasoning in science and in daily life is no mere academic exercise. It is liable to determine what you decide to do.

Logically, later information never invalidates earlier valid inferences. Logic is *monotonic*: With each additional premise, further conclusions follow validly from the premises. If a later premise negates an earlier conclusion, there is a contradiction, but a contradiction logically implies any conclusion whatsoever. Hence, logic never calls for the withdrawal of a conclusion. Some formulations of logic and some psychological theories of reasoning include the formal rule of *reductio ad absurdum* (e.g., Rips, 1994). According to this rule, if you make a supposition for the sake of argument, and can prove that it leads to a contradiction, then you are entitled to deny the supposition. When you validly infer a contradiction from premises alone, at least one premise is false, but the rule does not stipulate which is the offending premise. The moral is that logic allows you to detect inconsistencies and to use them to draw further consequences, but it never calls for you to withdraw a conclusion.

P. N. Johnson-Laird, Department of Psychology, Princeton University; Vittorio Girotto, Laboratoire de Psychologie Cognitive, Centre National de la Recherche Scientifique, Aix-en-Provence, France, and Department of Art and Design, Venice Architecture University, Venice, Italy; Paolo Legrenzi, Department of Art and Design, Venice Architecture University.

This research was supported in part by National Science Foundation Grant BCS 0076287 to P. N. Johnson-Laird to investigate strategies in reasoning and by a grant from the Italian Ministry of Universities and Scientific and Technological Research to Vittorio Girotto and Paolo Legrenzi. We thank many colleagues for help and advice: Tony Anderson, Victoria Bell, Ruth Byrne, Zachary Estes, Ino Flores d'Arcais, Yevgeniya Goldvarg, Michel Gonzalez, Uri Hasson, Hansjoerg Neth, Stefania Pizzini, Clare Walsh, and Yingrui Yang. We thank Susan Carey, Renee Elio, David Over, and Paul Thagard for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to P. N. Johnson-Laird, Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08544. E-mail: phil@princeton.edu

The retraction of conclusions calls for a special sort of reasoning. It is *nonmonotonic*; that is, you withdraw a previous conclusion in the light of subsequent information. In some cases, you have jumped to a conclusion on the basis of an assumption that you took to be true by default. Suppose, for example, that you took for granted that birds fly; you learned that Tweety is a bird; and so you inferred that Tweety flies. But then you learned that Tweety has a foot set in concrete. Obviously, you retracted your conclusion. As Ginsberg (1987) remarks about this example, "[t]he inference here is *non*monotonic. On learning a new fact. . .you were forced to retract your conclusion that he could fly" (p. 2). You can maintain your assumption that, at least in normal cases, birds fly. In other examples, such as the one about Paolo and the car, you have no option but to revise your beliefs, and logic cannot determine which premise you should retract.

The retraction of conclusions is pervasive in everyday life because events so often conspire to defeat inferences. To try to deal with such reasoning, researchers in artificial intelligence have developed various nonmonotonic systems (see, e.g., Brewka, Dix, & Konolige 1997) that allow for the withdrawal of old conclusions given new premises. In some of these systems, a premise such as that birds fly is treated as an idealization: By default, birds fly (e.g., Reiter, 1980). Hence, the conclusion that Tweety flies can be withdrawn if there is evidence to the contrary, but without the retraction of the default assumption that birds fly. Other nonmonotonic systems, however, allow for beliefs to be revised in the light of inconsistency (e.g., Doyle, 1979). Philosophers have also developed systems for the revision of beliefs in the face of inconsistencies (e.g., Gärdenfors, 1990; Harman, 1986; Levi, 1991). Psychologists have demonstrated the perseverance of social stereotypes in the face of conflicting evidence (see, e.g., Lepper, Ross, & Lau, 1986; Rehder & Hastie, 1996; Ross & Lepper, 1980). But the evidence on how naive individuals reason to consistency is sparse, apart from Revlis and his colleagues' pioneering studies (e.g., Revlis, 1974; Revlis, Lipkin, & Hayes, 1971), studies of the effects of the order of events on beliefs (e.g., Hogarth & Einhorn, 1992; Schlottmann & Anderson, 1995; Zhang, Johnson, & Wang, 1997), and some recent studies of the revision of beliefs (e.g., Dieussaert, Schaeken, De Neys, & d'Ydewalle, 2000; Elio, 1997; Elio & Pelletier, 1997; Politzer & Carles, 2001; Revlin, Cate, & Rouss, 2001). *Naive* refers to individuals who have not acquired an explicit mastery of formal logic or any cognate discipline. No one knows how such individuals reason from inconsistency to consistency.

The present article aims to overcome this deficit. The psychological problems to be solved are twofold. First, what are individuals computing when they reason to consistency? Second, how do they carry out these processes; that is, what are the underlying mental processes? In what follows, we attempt to answer both these questions. We present a model-based theory of how people reason to consistency and outline a computer program implementing the theory. We then assess this theory and other alternative accounts in the light of the experimental evidence. Finally, we draw some general conclusions.

The Model Theory of Reasoning to Consistency

The Computations in Reasoning to Consistency

What has to be computed when you reason to consistency? In our view, there are three main computations. First, you must detect an inconsistency within a set of propositions, typically a conflict between a conclusion that you have drawn and some evidence. Unless the evidence is dubious, or your inference is invalid, you have to withdraw your conclusion. Second, if your original propositions validly implied your conclusion, you must revise your belief in them. You must try to decide which of them to retract or to doubt. Third, in everyday life, you do not merely decide what propositions are dubious, but you also try to resolve the inconsistency. You aim to create an explanation of its origins. This process is important. Most previous studies of nonmonotonic reasoning and the revision of beliefs have tended to overlook the generation of explanations.

Consider again the case of Paolo and the car. As you sit waiting, you are concerned with what has happened to him. You think of various possibilities, and what you decide to do depends on which of them seems more likely. Of course, part of the process of generating possibilities may depend on your assessment of the premises: Did he really go to get the car, and if he did, would it really take only five minutes for him to return? But, your knowledge may yield possibilities directly, and these possibilities in turn may have further consequences for your belief in the premises. Your original conditional premise was, If Paolo has gone to get the car, then he will be back in five minutes. If you combine it, not with your original categorical premise, but with the fact that Paolo was not back in five minutes, then it follows validly that Paolo did not go to get the car. You could therefore retract this premise. Another possibility, however, is that the conditional premise itself is false. The two alternatives, coupled with your general knowledge, enable you to infer a variety of possibilities. One possible cause can lead in turn to further causal possibilities: For example, it is false that Paolo went to get the car (possibly he met a friend and went for a coffee), or it is false that if Paolo went to get the car then he will be back in five minutes (possibly it was stolen, or its engine would not start). The list of possibilities is indefinitely long. However, you are likely to eliminate probable causes, if you can, before you entertain improbable ones. The process of narrowing down the list may yield one overwhelmingly likely possibility, but often it will yield competing alternatives. Sometimes, it may fail to yield any possible explanation at all, which is what we refer to as the *Marie Celeste* phenomenon. The eponymous example is based on a true historical case: You believe that if you board a ship at sea, the crew will be there. You board the *Marie Celeste*. You discover that the crew is not aboard. But you are unable to infer what has happened to them. In sum, a major process in the resolution of inconsistency is the attempt to envisage a causal scenario—a diagnosis—that makes sense of the situation.

This analysis distinguishes three main processes for which we need an account. We summarize them in three questions about your performance in the case of Paolo and the car:

1. How do you detect an inconsistency? That is, how do you detect that the fact that Paolo has not returned conflicts with your beliefs?

2. Which propositions do you retract or come to doubt? Do you doubt that Paolo went to get the car, your conditional assumption that he would be back in five minutes, or both?

3. How do you generate explanations of the situation? How do you diagnose what may have happened to Paolo or to the car? This ability to envisage possible causes is at the heart of resolving the inconsistency. You need to make sense, if you can, of why Paolo has not returned in five minutes. You may have no definite answer, but only a view about what is likely or possible. You may have no idea at all.

Henceforth, we shall refer to the combination of these three processes as *reasoning to consistency*: the detection of an inconsistency, the revision of beliefs, and the explanation of the inconsistency. We describe the processes as though they occur in a sequence, but your attempt to make sense of the situation may itself determine which propositions you come to doubt.

The Model Theory

The theory that we present is based on the assumption that mental models play a central part in each of the three processes, and we refer to it as the *model theory*. We have argued elsewhere, as have others, that deductive reasoning depends on understanding the meaning of premises and using this meaning and general knowledge to construct a set of mental models of what the premises describe (e.g., Johnson-Laird, 1993; Johnson-Laird & Byrne, 1991; Polk & Newell, 1995). A mental model is, by definition, a representation of a possibility. Its structure and content capture what is common to the different ways in which the possibility might occur (Barwise, 1993). Hence, a central component of reasoning is the generation of possibilities. A conclusion is *necessary* if it holds in all the models of the premises, and it is *possible* if it holds in at least one model of the premises; its *probability* depends on the proportion of equiprobable models in which it holds or on numerical probabilities attached to models (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). Insofar as possible, mental models are iconic; that is, their parts correspond to the parts of what they represent, and their structures correspond to the structures of what they represent (see, e.g., Peirce, 1931–1958, Vol. 4, paragraph 447). Visual images are iconic too, and mental models may be experienced as visual images. However, you can have a mental model of an abstract proposition, such as The President is not allowed to own a house. You may form images of the President and a house, but no mere image can capture the meaning of negation, permission, or ownership. One advantage of the iconic nature of mental models—an advantage that Peirce exploited in his own diagrammatic system for logic—is that you can use some propositions to build a model and then use the model to draw an emergent conclusion that does not correspond to any of these propositions.

Suppose that the following proposition expresses an exclusive disjunction in which only one of the two clauses is true:

There is not a circle or else there is a triangle.

Sentences are normally used to express propositions, and for convenience we use the standard term in logic, *proposition*, to refer to this use of sentences. We use another standard term, an *atomic* proposition, to refer to a proposition that contains neither negation nor any connectives. For example, the preceding assertion contains two atoms: there is a circle, there is a triangle. The model theory depends on a key assumption about the representation of propositions:

1. *The principle of truth:* Individuals represent propositions by constructing sets of mental models in which each model represents a true possibility Each clause in a proposition, whether it is an atomic proposition or the negation of an atomic proposition, is represented in a mental model only if it is true in that possibility.

The principle of truth is subtle, and the easiest way to grasp it is to consider an illustrative example. The mental models of the exclusive disjunction, There is not a circle or else there is a triangle, represent only the two true possibilities, and within them, they represent the two clauses in the disjunction (there is not a circle, there is a triangle) only when they are true within a possibility. We depict these two mental models in the following diagram, in which each row denotes a separate model of a possibility:

```
¬o
        △
```

where ¬ denotes negation, o denotes a model of the presence of a circle, and △ denotes a model of the presence of a triangle. Hence, the first model does not represent explicitly that it is false that there is a triangle in this possibility; and the second model does not represent explicitly that it is false that there is not a circle in this possibility, that is, there *is* a circle. Reasoners make "mental footnotes" to keep track of the information about what is false, but they soon forget these footnotes. If they do keep track of the footnotes, however, then they can use them to flesh out their mental models into *fully explicit* models and thereby overcome the principle of truth. The mental models of the preceding exclusive disjunction can be fleshed out to yield the following fully explicit models:

```
¬o     ¬△
 o      △
```

These models also correspond to the fully explicit models of the biconditional proposition, If and only if there isn't a circle then there isn't a triangle. But, when most people encounter the disjunction, they do not grasp the equivalence, because they rely only on mental models.

The mental models of conditionals are rudimentary. For a conditional such as, If there is a circle then there is a triangle, the mental models represent explicitly the possibility in which the antecedent subordinate clause (there is a circle) is true, whereas the possibilities in which it is false are represented by a wholly implicit model (shown here as an ellipsis):

```
o    △
  . . .
```

Individuals need to make a mental footnote that the antecedent is false in the possibilities that the implicit model represents. If they retain this footnote, they can flesh out the mental models into fully explicit models:

o △
¬o △
¬o ¬△

The mental models of a biconditional, If and only if there is a circle then there is a triangle, are identical to those for the conditional above. All that differs is that the mental footnote indicates that both the antecedent and the consequent are false in the possibilities represented by the implicit model. Table 1 summarizes the mental models and the fully explicit models of the basic set of sentential connectives. These interpretations, however, are affected by both semantic and pragmatic factors (see, e.g., Evans & Over, 1996; Garnham & Oakhill, 1994). Johnson-Laird and Byrne (2002) have described the process by which these factors can modulate the interpretation of conditionals. They can flesh out a model with further information, but they can also prevent the construction of a model of a possibility.

With experience, reasoners develop a variety of model-based strategies. Van der Henst, Yang, and Johnson-Laird (2002) have described these strategies, but we forgo the details here. Readers who have not encountered the theory before may worry about such concepts as mental footnotes and implicit models. In fact, the theory postulates that individuals normally reason using mental models, but that in simple inferences they can flesh out their models to make them fully explicit. The comprehension of the premises yields a set of mental models; the evaluation of a given conclusion is a process of verifying it in relation to the set; and the formulation of a conclusion is a process of describing the set.

There are several "tell-tale" signs of the use of mental models. One sign is that reasoners can generate counterexamples to refute invalid inferences. Consider the following problem:

More than half the people in the room speak French.
More than half the people in the room speak Italian.
Does it follow that more than half the people in the room speak French and Italian?

Individuals routinely refute the conclusion by envisaging a possibility in which the premises are true but the conclusion is false; for example, there are five people in the room, three speak each language, but only one speaks both languages (Bucciarelli & Johnson-Laird, 1999; Johnson-Laird & Hasson, 2003). A second sign of models is that inferences that call for multiple models take longer and are more error prone than those that call for only a single model (e.g., Byrne & Johnson-Laird, 1989; Schaeken, Johnson-Laird, & d'Ydewalle, 1996). A third sign is that erroneous conclusions correspond to some of the models of the premises, typically just a single model (e.g., Bara, Bucciarelli, & Johnson-Laird, 1995; Bauer & Johnson-Laird, 1993). Ormerod, Manktelow, and Jones (1993) have corroborated a *minimal completion* hypothesis, according to which reasoners construct models of only what is minimally necessary. To explain tasks going beyond straightforward deduction, it is necessary to make additional assumptions (see Girotto & Gonzalez, 2001; Legrenzi, Girotto, & Johnson-Laird, 1993; Johnson-Laird et al., 1999). We likewise make some additional assumptions to explain reasoning to consistency.

Table 1
Mental Models and Fully Explicit Models for the Main Sentential Connectives

Connective	Mental models		Fully explicit models	
A and B	A	B	A	B
A or else B	A		A	¬ B
		B	¬ A	B
A or B, or both	A		A	¬ B
		B	¬ A	B
	A	B	A	B
If A then B	A	B	A	B
	. . .		¬ A	B
			¬ A	¬ B
If and only if A then B	A	B	A	B
	. . .		¬ A	¬ B

Note. "¬" denotes negation, and ". . ." denotes a wholly implicit model. Each row represents a model of a possibility.

The Detection of Inconsistencies

The first process in reasoning to consistency is the detection of an inconsistency among a set of propositions. It is tempting to think of inconsistency as a conflict between just two propositions, one of the form *A*, and the other either its contrary or its contradiction: *Not-A*. Unfortunately, inconsistency can occur in a set of propositions in which any proper subset is consistent. For example, consider a set of propositions based on three atoms, *A*, *B*, and *C*:

A or B, or both.
Not-B or C, or both.
Not-A and not-C.

Each pair of propositions is consistent, but the three together are inconsistent. In general, the detection of inconsistency is intractable (technically, it is *NP-complete*, Cook, 1971). That is, it makes bigger and bigger demands on time and memory as the number of distinct atoms in a set of propositions increases. These demands can increase so that no feasible computational system could yield a result, not even a computer as big as the universe running at the speed of light. A set of, say, 100 atomic propositions allows for 2^{100} possibilities, because each atom can be either true or false. This number is vast, and, in the worst case, a test of consistency calls for check of every possibility. If one possibility could be checked in a millionth of a second, it would still take over 40 thousand million million years to examine them all. Of course, the intractability of a domain does not mean that every problem within it is impossible to solve. Small-scale problems of the sort that we have investigated are solvable both psychologically and computationally.

How do people decide whether or not a set of propositions is consistent? Prior to our research, psychologists do not seem to have addressed the question. One method, however, can be based on formal rules of inference, such as

If A then B.
A.
Therefore, B.

where *A* and *B* can refer to any propositions whatsoever. Formal rules of this sort have been used to explain deductive inferences

from premises to conclusions (e.g., Braine & O'Brien, 1998; Rips, 1994). But the evaluation of consistency is a different task. You need to determine whether a set of propositions can all be true. Nevertheless, formal rule theories can be adapted to cope with consistency. You select a proposition from the set, and try to prove its negation from the remaining propositions. If you succeed, then the set is inconsistent; otherwise, it is consistent. The procedure seems implausible psychologically, and so we propose a different theory based on mental models.

A single assumption extends the model theory to deal with the evaluation of consistency:

> 2. *The principle of modeling consistency:* Naive individuals evaluate the consistency of a set of propositions by searching for a single mental model of a possibility that satisfies all the propositions. If there is such a model, then the set is consistent; otherwise, it is inconsistent. The more models that have to be examined in the search, the harder the task will be.

We have written a computer program that implements the model theory of all three steps in reasoning to consistency: the detection of inconsistency, the revision of beliefs, and the creation of causal explanations to resolve inconsistencies.[1] The first stage of the program uses the principle of modeling consistency to evaluate whether or not a set of propositions is consistent. It searches for a single model that satisfies every proposition in the set; that is, each proposition is true in the model. The overall structure of this stage of the program is as follows: First, it takes as input a set of propositions. Second, it constructs their models. Third, it searches for a model that satisfies all the propositions. If it succeeds, it returns the result that the propositions are consistent; otherwise, it returns the result that they are inconsistent. The program operates at two levels of expertise. At its simple level, it carries out these processes on mental models without footnotes. At its advanced level, it uses mental footnotes to flesh out mental models into fully explicit models, and therefore makes no errors. For any given problem, it produces an output at both levels. We now describe each of the processes in the first stage of the program in more detail.

First, the input to the program is a set of propositions, which can contain negation and sentential connectives, such as "if," "or," and "and," for example,

There is an ace ore not comma there is a king and there is a queen.
There is an ace and not there is a king.

The program treats each affirmative clause as an atomic proposition: *Ore* denotes an exclusive disjunction (see Table 1); *not* denotes negation, which precedes the clause it negates; and "comma" is equivalent to a left parenthesis. Hence, the parser treats these two propositions as having the form

Ace ore not (king and queen).
Ace and not king.

Second, the program parses each sentence to produce the models of the proposition it expresses. The program has a lexicon that gives the interpretations of connectives presented in Table 1. It uses a compositional semantics to construct models as it parses sentences according to a grammar. Because the clauses interrelated with a sentential connective can themselves contain connectives, the program constructs models recursively. At the heart of this process are the compositional procedures for negation and for making conjunctions of models: In this way, it can cope with all the connectives in Table 1. The negation of a set of models is its complement from the set of all possible models based on the same atoms. For example, consider an exclusive disjunction of the form

Not ace ore king.

Its fully explicit models are as follows:

¬ Ace	¬ King
Ace	King

To negate this set, the program recovers the list of every atom that occurs in the set (Ace King). It constructs the set of all possible models containing these atoms:

Ace	King
Ace	¬ King
¬ Ace	King
¬ Ace	¬ King

It then returns every model in this set that is not in the models for the disjunction

Ace	¬ King
¬ Ace	King

The procedure depends on fully explicit models, so, to negate a set of models, individuals need to flesh out their mental models into fully explicit models. Because this task is difficult, naive individuals are unable to envisage the possibilities in which all but the simplest propositions are false (Barres & Johnson-Laird, 2003). The program's procedures for forming conjunctions of pairs of mental models and of pairs of fully explicit models are listed below.

1. The conjunction of a pair of implicit models yields the implicit model: . . . and . . . yield . . .

2. The conjunction of an implicit model with a model representing propositions yields the null model (akin to the empty set) by default: for example, . . . and B C yield nil. But, if none of the atomic propositions (B C) is represented in the set of models containing the implicit model, then the conjunction yields the model of the propositions: for example, . . . and B C yield B C.

3. The conjunction of a pair of models containing respec-

[1] The source code in Common Lisp is on www.princeton.edu/~psych/PsychSite/~phil.html.

tively a proposition and its negation yields the null model: for example, A ¬ B and ¬ A yield nil.

4. The conjunction of a pair of mental models in which a proposition, B, in one model is not represented in the other model depends on the set of models of which this other model is a member. If B occurs in at least one of these models, then its absence in the current model is treated as negation: for example, A B and A yields nil. But, if B does not occur in one of these models, for example, only its negation occurs in them, then its absence is treated as equivalent to its affirmation, and the conjunction (following the next procedure) is A B and A yields A B.

5. The conjunction of a pair of fully explicit models free from contradiction updates the second model with all the new propositions from the first model: for example, ¬ A B and ¬ A C yield ¬ A B C.

Note that only mental models may be implicit and therefore call for the first two procedures. The mental models of the proposition, Ace ore not comma king and queen, are

Ace

¬ King	Queen
King	¬ Queen
¬ King	¬ Queen

The mental model of the proposition, Ace and not king, is

Ace ¬ King

Third, the program searches for a model that holds for all the propositions. It starts with the first model of the first proposition and searches for a consistent interpretation with the second proposition. In general, it loops through each model of the second proposition, forming a conjunction with the model of the first proposition by using the procedures shown above. As soon as the conjunction is viable, that is, it does not return the null model, it proceeds to the next proposition, and so on. However, if the conjunction yields the null model, it tries the next model of the second proposition. When it has exhausted all the models of the second proposition, it tries the next model of the first proposition, and so on. If two models are inconsistent, then their conjunction returns the null model (see Procedure 3 above). But in the example, the absence of a proposition is treated as equivalent to its negation (see Procedure 4 above). Hence, the conjunction of the first mental model of the first proposition with the first model of the second proposition yields a consistent interpretation:

Ace ¬ King

and so the program yields the result that the propositions are consistent. At its advanced level of performance, the program uses fully explicit models; that is, it uses footnotes to flesh out mental models into fully explicit models, which represent both what is true and what is false in a possibility. The fully explicit models of the first proposition in the example are as follows:

Ace	King	Queen
¬ Ace	King	¬ Queen
¬ Ace	¬ King	Queen
¬ Ace	¬ King	¬ Queen

In this case, the program considers all pairwise conjunctions of these models with the model of the second proposition. They all yield the null model. Hence, contrary to the response that the program makes when it uses mental models, the two propositions are, in fact, inconsistent. Appendix A presents the verbatim output of the first stage of the program to this problem.

Because human working memory has a limited processing capacity, the model theory predicts that the greater the number of models that have to be constructed, the harder the task should be. It should be difficult to have to backtrack and to consider an alternative model of an earlier proposition. But, as the computer program revealed in its output shown in Appendix A, the theory also yields a more surprising prediction: Illusions of consistency should occur if individuals rely on mental models as opposed to performing at an advanced level. That is, in certain quite simple cases, individuals should judge that a set of propositions is consistent when in fact the set is inconsistent. The program also predicts illusions of inconsistency.

The Revision of Propositions

What happens when you discover an inconsistency among a set of propositions? If there is no proposition among them that you know with a greater certainty than the others, then you may choose to defer any attempt to reach consistency until you have more information. But what happens when the inconsistency arises from incontrovertible facts? There are a variety of procedures that reasoners might try, depending on the circumstances. If the inconsistency arises from a consequence inferred from the propositions, then their first step is likely to be to check whether the inference is valid. As Wason (1964) demonstrated, an inconsistency generated from an invalid inference is likely to lead reasoners to think again about the inference. If an inconsistency arises from a conflict between a fact and a valid inference from propositions, then at least one of the propositions must be given up. But which one? If individuals have made an arbitrary assumption or an assumption by default, then they can give up such an assumption (see Johnson-Laird & Byrne, 1991, for an account of how the model theory deals with such cases). But what happens when none of the relevant beliefs is an arbitrary or default assumption? The simplest case occurs when the facts conflict with just a single proposition. For example, you put some milk in the fridge, and so you believe that there is milk in the fridge. A short while later, you go to the fridge to get some milk, and you discover that there is none. Naturally, you cease to believe that there is milk in the fridge. A more complex conflict occurs when there is no single proposition with which the facts conflict, though they conflict with a set of propositions as a whole. Consider the following case:

Evelyn tells you, If Nicola did the shopping, then there's milk in the fridge.
Vivien tells you, Nicola did the shopping.
But you discover that there is no milk in the fridge.

You may know that one of your two informants is misinformed, unreliable, or worse. You may know that Nicola is unlikely to do the shopping or else likely to forget to buy milk. You are likely to retract whichever proposition, if any, that your knowledge undermines.

The cases that chiefly concern us are subtler. They occur when all the propositions are equally plausible but collectively conflict with the facts of the matter. What propositions are individuals likely to retract or doubt in this case? The model theory implies that individuals should be susceptible to illusory thinking in this case too. When a proposition has mental models conflicting with the unequivocal facts, the proposition should seem to conflict with the facts. There may be no real conflict, but the compatibility of the proposition depends on possibilities that are not represented in its mental models, but only in its fully explicit models. Individuals usually overlook these models, and so they should be likely to doubt the proposition.

Suppose that no proposition in the set is in such an apparent conflict with the facts; what then? The model theory yields another possibility. If there is just one proposition with mental models that fails to represent the facts, then it too may seem to conflict with the facts. All the other propositions, by definition, represent the facts, and so this proposition is unique in failing to represent it. Hence, reasoners should be likely to doubt this proposition. Evans and his colleagues have shown that the matching of clauses can have important consequences for reasoning (for a review, see Evans, Newstead, & Byrne, 1993), but the mismatches that concern us are between models. And, as we will show, they do not necessarily correspond to those between clauses. When an inconsistency occurs between a fact and a single proposition, individuals retract the proposition. Otherwise, the theory postulates the following assumption:

> 3. *The mismatch principle:* When an inconsistency occurs between facts and a set of propositions that are all equally plausible, individuals retract a proposition that has mental models conflicting with the facts and none matching them; otherwise, they retract a proposition that fails to represent the facts.

When neither of these clauses applies, the theory does not predict any bias with regard to which propositions reasoners are likely to abandon.

As an illustration, consider how the program implementing the theory operates during its second stage, which is based on the mismatch principle. This second stage of the program is called only if its first stage, which we described earlier, detects an inconsistency. The second stage determines which proposition, if any, to reject because it mismatches the facts, and it operates with mental models and with fully explicit models. Its overall structure is as follows: First, the input to the stage is the set of models for each proposition and the model of the facts. Second, the program compares the model of the facts with the models of each of the propositions according to the mismatch principle. Third, if only one proposition mismatches the facts, then this proposition is rejected; otherwise, the program does not reject any proposition. Finally, the program constructs a revised model of the facts incorporating any proposition that it has not rejected, and it constructs models of counterfactual possibilities based on the proposition that it has rejected.

As an example, we describe how this stage of the program treats an inconsistency with a modus ponens inference:

You believe, If the plane is on course, then the radar should show water.

You believe, The plane is on course.

You learn for a fact, The radar does not show water.

The first stage of the program, as we have seen, evaluates the consistency of the propositions and, with both mental models and fully explicit models, detects the inconsistency between the two initial propositions and the facts in the final proposition. The second stage has an input of the mental models of the conditional proposition, shown in abbreviated form:

plane-on-course radar-shows-water
. . .

the mental model of the categorical proposition

plane-on-course

and the mental model of the fact

¬ radar-shows-water

The program returns one of three possible outcomes for each proposition: The model of the facts occurs in a model of the proposition (i.e., there is at least one match); the model of the facts does not occur in any model of the proposition, but conflicts with at least one model (i.e., there is a mismatch); or neither of the two previous cases holds (i.e., the proposition does not refer to the facts). In the example, the program discovers that the model of the facts conflicts with the explicit mental model of the conditional proposition, because their conjunction yields the null model. Hence, according to the mismatch principle, it rejects the conditional proposition. With fully explicit models, however, the model of the facts matches one of the models of the conditional, but it is not represented in the model of the categorical proposition. Hence, according to the mismatch principle, the program rejects the categorical proposition. Appendix B presents the verbatim output of the program for this example, and it shows the revised model of the facts and the models of the counterfactual possibilities.

Biconditionals yield only two possibilities, and reasoners are more likely to consider their fully explicit models than the three fully explicit models that ordinary conditionals yield. One way to enhance naive individuals' deductive performance is therefore to use biconditional propositions instead of ordinary conditionals (see, e.g., Johnson-Laird & Byrne, 1991). Given a biconditional, such as

If and only if the plane is on course then the radar shows water

the program, at its advanced level, constructs two fully explicit models:

plane-on-course	radar-shows-water
¬ plane-on-course	¬ radar-shows-water

In a conflict with modus ponens, the program now detects that the model of the fact, ¬ radar-shows-water, matches the second of these models. It therefore rejects the categorical proposition, which fails to represent the fact. The program accordingly predicts that expertise should influence which propositions individuals reject. Readers may wonder whether mismatches might occur solely as conflicts between the surface clauses of sentences (see Elio & Pelletier, 1997). This hypothesis makes predictions that differ from those of the model theory. We return to this point in the section assessing the theory.

The Explanation of Inconsistencies

The ability to explain inconsistencies transcends the revision of beliefs. To revert to our initial example, when Paolo fails to return in the car, you do not merely cease to believe that he went to get the car or your conditional belief that if he did, he will be back in five minutes. You try to envisage what is likely to have happened to him. The third process in reasoning to consistency is accordingly to create a diagnostic explanation that resolves the inconsistency. People are able to generate such explanations. This skill, which some philosophers refer to as *abduction* (Peirce, 1903/1955), seems unremarkable, but no existing computer program comes close to matching human ability. Abduction is a species of induction in that its results may be false even if its premises are true, but it goes beyond mere generalization into the domain of causality. For example, you might explain Paolo's absence by inferring that he ran into a complicated one-way system and so is taking a long time to return. In terms of what is computed, our principal claims are threefold. First, causal explanations can be decomposed into temporally ordered possibilities; second, individuals use their general knowledge of such possibilities to construct a causal chain that explains the inconsistent fact; and, third, the mismatch principle biases the nature of explanations. We amplify each of these points in turn.

The knowledge that is most pertinent to explaining everyday inconsistencies is knowledge of causes and effects. Individuals are unlikely to attribute inconsistencies to events that have no causes, though these sorts of explanation could occur in some cultures (see, e.g., Morris, Nisbett, & Peng, 1995). In fact, no culture, as far as we know, eschews causal explanations, and many cultures, including our own, put the highest value on them. A plausible explanation should be a causal one, and, given a choice between a deduction and a causal explanation, individuals in our culture should be biased toward the latter.

According to the model theory, the meaning of a causal relation between two states of affairs, *A* and *B*, concerns what is possible and what is impossible in their co-occurrences. The claim is controversial, but it has been corroborated experimentally (Goldvarg & Johnson-Laird, 2001). We emphasize that the theory concerns the meanings of causal relations, not how these relations are induced from observations. In daily life, the normal constraint on a causal relation between *A* and *B* is that *B* does not precede *A* in time (see, e.g., Tversky & Kahneman, 1982). Hence, the theory adopts this constraint.

According to the theory, a proposition of the form, *A will cause B*, such as

Pulling the trigger will cause the gun to fire

means that only certain events are possible. The most salient possibility is one in which both the cause and the effect occur, and the effect does not precede the cause. Other possibilities in which the cause does not occur are compatible with the proposition, but the proposition rules out as impossible the case in which the trigger is pulled but the gun does not fire. Hence, the proposition is compatible with three temporally ordered possibilities:

pull trigger	gun fires
¬ pull trigger	gun fires
¬ pull trigger	¬ gun fires

Ordinary causes are thus sufficient to bring about their effects, but not necessary for these effects to occur, because the effects may have other causes; for example, if the trigger is unguarded and the gun is dropped, it may fire. Some causes, however, are unique; for example, an extreme deficiency of vitamin C is the unique cause of scurvy. In this case, there are only two possibilities:

vitamin C deficiency	scurvy
¬ vitamin C deficiency	¬ scurvy

Unique causes are both necessary and sufficient for their effects.

In addition to causes, there are causal relations that concern enabling states of affairs, for example, "Exercise allows you to grow stronger." If you are ill, exercise may fail to increase your strength; likewise, you may grow stronger even if you don't exercise, for example, by adopting a special diet. Ordinary enabling conditions are therefore compatible with all four temporally ordered contingencies:

exercise	grow stronger
exercise	¬ grow stronger
¬ exercise	grow stronger
¬ exercise	¬ grow stronger

Some enabling conditions, however, are unique, for example, "Oxygen allows life to develop." They rule out the case in which the effect occurs without the enabler: Without oxygen, there is no life. Hence, the proposition is compatible with only three temporally ordered possibilities:

oxygen	life
oxygen	¬ life
¬ oxygen	¬ life

Unique enabling conditions are thus necessary to bring about effects, but not sufficient to bring them about.

The sets of possibilities distinguish between the meaning of a causal relation: *A will cause B*, and an enabling relation: *A will allow B*. As a consequence, the logical implications of the two sorts of proposition should also differ. This claim is controversial in two ways. On the one hand, probabilistic theories of causation cannot readily distinguish between causes and enabling conditions,

because both of them increase the probability of the effect (pace, e.g., Cheng, 1997; Cheng & Novick, 1990; Reichenbach, 1956; Suppes, 1970). The main evidence for a probabilistic semantics is that people judge that a causal relation holds in cases in which the antecedent is neither necessary nor sufficient to bring about the effect (e.g., Cheng & Novick, 1990; Cummins, Lubart, Alksnis, & Rist, 1991). One might therefore suppose that causal relations are intrinsically probabilistic. Certainly, people often induce causal relations from probabilistic data. Yet, it does not follow that the meaning of causal relations is probabilistic. The present hypothesis is that the meaning of a causal relation is not probabilistic, though the evidence supporting the relation may be probabilistic.

On the other hand, current psychological theories argue that causes and enabling conditions differ, but the difference is not in their meaning or logic. The argument is based on Mill (1843/1874). It first influenced philosophers, then jurists and psychologists. They have proposed many candidate distinctions:

1. Causes are recent, whereas enablers are earlier (Mill, 1843/1874).

2. Causes are abnormal or rare events, whereas enabling conditions are normal or common (e.g., Hart & Honoré, 1985).

3. Causes are inconstant, whereas enablers are constant (Cheng & Novick, 1991).

4. Causes violate a norm, whereas enablers do not (e.g., Einhorn & Hogarth, 1986; Kahneman & Miller, 1986).

5. Causes are conversationally relevant in explanations, whereas enablers are not (e.g., Hilton & Erb, 1996; Turnbull & Slugoski, 1988).

and so on. All these distinctions may be true, but the model theory postulates that causal and enabling relations differ in their meaning too. That is, they refer to different sets of possibilities.

What is the best sort of explanation to resolve an inconsistency between facts and a set of propositions? As we have argued, it should be a causal chain, but how long should the optimal chain be? As an example, consider the following problem:

If someone pulled the trigger, then the gun fired.
Someone pulled the trigger, but the gun did not fire.
Why not?

A minimal explanation would be:

There were no bullets in the chamber.

It also accords with the mismatch principle because it rules out the conditional, as opposed to the categorical, premise. But, such an explanation is ad hoc and unmotivated: It in turn stands in need of an explanation. A more plausible explanation provides such motivation in terms of a cause:

A prudent person unloaded the gun and there were no bullets in the chamber.

Of course, one might also ask for an explanation of this cause, but the longer a causal chain, the more improbable it becomes. Strictly speaking, the preceding explanation is already more improbable than the minimal explanation that there were no bullets in the chamber. But, according to the present theory, reference to both a cause and an effect is optimal. The cause accounts for the effect that resolves the inconsistency, but the sequence is not so long that it seems improbable.

The hypothesis that cause and effect provide an optimal explanation violates a common theoretical assumption about the revision of beliefs. Philosophers have long argued that changes to beliefs in the face of an inconsistent fact should be as conservative as possible, so that the accommodation of the new fact is accompanied by a minimal change to other beliefs. As James (1907) wrote, "[The new fact] preserves the older stock of truths with a minimum of modification, stretching them just enough to make them admit the novelty" (p. 59). Recent theorists have also advocated this principle (e.g., Gärdenfors, 1988; Harman, 1986). But, if the model theory is correct, then the principle is wrong, because an explanation that posits both a cause and an effect is less minimal than an explanation that posits only an effect.

According to the theory, a cause and its effect should be more convincing than the effect alone. But, the cause and its effect should also be more convincing than the cause alone, which leaves some uncertainty about whether the gun remained unloaded until the trigger was pulled. The theory further predicts that the cause alone should be more convincing than the effect alone. The models of the cause-and-effect relation

unload	¬ bullets
¬ unload	bullets
¬ unload	¬ bullets

make it easy to infer the effect (no bullets) from the cause (unload), because, given the cause, the effect is the only possibility. In contrast, it should be harder to infer the cause (unload) from the effect (no bullets), because there is no unique cause in the preceding models; for example, all the bullets may have been fired. Of course, unique causes are exceptions to this prediction (cf. Cummins et al., 1991; Markovits, 1984), but the prediction should hold in general.

The mismatch principle yields a further prediction about problems based on biconditionals, such as

If and only if someone pulled the trigger, then the gun fired.

In this case, as we saw earlier, the mismatch principle predicts that individuals should be less biased to abandon the biconditional given an inconsistency with a modus ponens inference. Hence, they should show a concomitant shift toward a preference for explanations that abandon the categorical premise.

How do individuals create causal explanations? Complete causal explanations are unlikely to be sitting in long-term memory waiting to be elicited by relevant problems; rather, mental processes construct them from more elementary causal relations represented in knowledge. The model theory postulates that individuals know about many causal and enabling relations, which are represented in knowledge as explicit models of sets of possibilities. Two special processes occur.

First, the information in a scenario yielding an inconsistency can trigger a particular possibility from a set of explicit models in knowledge, and in this case, if the explicit model in knowledge is inconsistent with a model of the scenario, then by default the model in knowledge takes precedence. This process also occurs in the interpretation of propositions. It can thereby modulate the normal meaning of a sentential connective. For example, the conditional:

If Pat is not in Rio then she is in Brazil

is compatible with only two possibilities, and Pat is in Brazil in both of them (for supporting evidence, see Johnson-Laird & Byrne, 2002). The knowledge that Rio is in Brazil blocks the construction of a model representing the possibility that Pat is in Rio but not in Brazil. Individuals ordinarily take this possibility—in which both the antecedent and the consequent of a conditional are false—to be compatible with a conditional.

Second, the model of a possibility that is triggered in general knowledge can in turn trigger a further model in another set in knowledge, with the result that the process yields a causal chain resolving the inconsistency. Such chains are highly likely to be novel, in the sense that the individual has never thought of them before. They are created rather than constructed by rote. We integrate the preceding account in the following principle:

> 4. *The principle of causal knowledge:* Different models of temporally ordered possibilities represent knowledge of causes and enabling conditions. These possibilities can be used to construct causal chains. An optimal chain consists of a cause and its effect. In the resolution of an inconsistency, such a chain takes precedence over the models of the propositions and explains the inconsistent fact.

The third stage of the computer program modeling the theory implements this principle. Its overall structure, which operates with mental models and with fully explicit models, is as follows: First, the input to the program is the revised model of the facts constructed by the second stage of the program, which incorporates the remaining premises into the model of the facts. Second, the program uses the content of this model to search its knowledge base, and, if possible, to trigger a possibility relevant to accounting for the facts. If the search is successful, this possibility is used to modulate the model of the facts. Third, given the new model, the program uses it to search its knowledge base again, and, if possible, to trigger a possibility that causes the new effect that the model contains. In this way, the program constructs a causal chain to resolve the inconsistency.

We can illustrate both the program and the principle of causal knowledge using the example

If someone pulled the trigger then the gun fired.
Someone pulled the trigger, but the gun did not fire.
Why not?

The first stage of the program detects the inconsistency, and the second stage uses the mismatch principle to abandon the conditional, because the mental model of the facts conflicts with the explicit mental model of the conditional above. The program conjoins the model of the remaining categorical proposition with the model of the facts to yield a revised model of the facts:

pulled-trigger ¬ gun-fired

This model is the input to the third stage of the program. It triggers several possibilities in the program's knowledge base, and the program makes an arbitrary choice among them:

gun-broken ¬ gun-fired

which represents the possibility that the gun is broken and does not fire. The program does not construct multiple models of possibilities as it builds an explanation, but instead constructs a model of single possibility:

pulled-trigger gun-broken ¬ gun-fired

The conjecture that the gun is broken is unmotivated. According to the theory, it stands in need of explanation. The program constructs such a cause by using the preceding model to search its knowledge base for a possibility that accounts for the breaking of the gun. Among its knowledge is the notion that if you drop a gun, then it may break:

gun-dropped gun-broken

The result is a causal chain of the sort that individuals construct for themselves in order to resolve inconsistencies. The explanations are not prestored in the program's knowledge base; rather, their component possibilities are retrieved from it, and then fitted together to make explanations.

With fully explicit models, the program uses the mismatch principle to abandon the categorical proposition, because its model does not represent the fact. It constructs a model of the current facts:

¬ pulled-trigger ¬ gun-fired

These facts trigger the possible explanation in the knowledge base that the person did not have enough strength to pull the trigger:

¬ enough-strength

This possibility, in turn, triggers the putative cause that the person has a partial paralysis:

partial-paralysis

This model completes the causal chain to resolve the inconsistency: Partial paralysis caused the individual to lack enough strength to pull the trigger.

Appendix C shows a complete output of the program given the present problem. The model of the facts can trigger several possibilities in the knowledge base. The program then makes an arbitrary choice among them. Individuals, however, are likely to use further knowledge to try to choose the most likely possibility.

The model theory and its computer implementation make testable predictions about all three processes in reasoning to consis-

tency. When individuals evaluate a set of propositions as consistent, the task should be easier when the initial model of the proposition suffices than when they have to search for an alternative model of the propositions. They should also succumb to illusions of consistency and illusions of inconsistency. When they have detected an inconsistency, they should be biased to reject whichever proposition, if any, yields models that mismatch the model of the fact. And to resolve the inconsistency, they should construct a causal chain made up from component possibilities in general knowledge—a chain that goes beyond a minimal explanation to include both a cause and an effect. The next section of the article assesses the model theory and other alternative accounts in the light of experimental tests of these predictions.

An Assessment of the Theory

The Detection of Inconsistencies

Until recently, there has been a dearth of studies of how individuals detect inconsistencies. The principle of modeling consistency predicts that the more models reasoners have to consider, the harder the task should be. Several experiments have corroborated this prediction. In one study, the participants had to state whether or not a set of propositions could all be true at the same time, which is simpler for them to understand than a direct request for a judgment of consistency (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000). The results showed that the task was easier with problems based on conditionals, such as

If there isn't an apple then there is a banana.
If there is a banana then there is a cherry.
There isn't an apple and there is a cherry.

than with logically equivalent problems based on disjunctions:

There is an apple or there is a banana, or both.
There isn't a banana or there is a cherry, or both.
There isn't an apple and there is a cherry.

The equivalence between the two sets of assertions also follows from empirical studies in which the participants list what is possible given the assertions or infer one sort of assertion from the other (e.g., Johnson-Laird & Byrne, 2002). As the computer program shows, to judge the consistency of the problem based on conditionals calls only for the construction of a single initial model of the two conditionals:

¬ apple banana cherry

To judge the consistency of the problem based on disjunctions, however, calls for the construction of multiple models. The first proposition yields the following possibility:

apple

and the absence of a banana in this model is consistent with the second proposition:

apple ¬ banana

However, the model is not consistent with the third proposition. Hence, as the program shows, individuals have to construct an alternative model:

¬ apple banana cherry

This possibility is compatible with all three propositions. But the need to search for an alternative possibility does increase the difficulty of the problem.

In a second study, the participants had to describe possibilities consistent with sets of propositions (Legrenzi, Girotto, & Johnson-Laird, 2003). If they thought that there was no such possibility, that is, that the set was inconsistent, they had to respond that the task was impossible. For example, consider a problem of the following form:

The chair is saleable if and only if it is elegant.
The chair is elegant if and only if it is stable.
The chair is saleable or it is stable, or both.

The program begins by constructing a model of the chair satisfying the first two propositions:

saleable elegant stable

This model is compatible with the truth of the third proposition, and so the description of the chair is

saleable, elegant, and stable.

In contrast, consider the following problem:

The chair is unsaleable if and only if it is inelegant.
The chair is inelegant if and only if it is unstable.
The chair is saleable or it is stable, or both.

where the propositions were based on different implicit Italian negatives, such as "invendible" (unsaleable):

¬ saleable ¬ elegant ¬ stable

But this model conflicts with the third proposition. Hence, reasoners would have to construct an alternative model of the first two propositions. A model of the following possibility is compatible with all three propositions:

saleable elegant stable

Of course, the second problem might be harder because it contains implicit negatives, and so the experiment counterbalanced their occurrence in the two sorts of problem (i.e., one-model and alternative-model problems). The results showed that the participants were more accurate with the one-model problems (overall 97% correct) than with the alternative-model problems (overall 39% correct).

Although the model theory predicts the phenomena, they may be open to other explanations. Theories based on formal rules might be framed in some way to accommodate the results. But one phenomenon is at present predicted only by the model theory and

its principle of truth: the occurrence of illusory inferences. As the computer program shows, two sorts of illusion should occur. In illusions of *consistency*, reasoners should infer that a set of propositions is consistent when, in fact, it is inconsistent; and in illusions of *inconsistency*, reasoners should infer that a set of propositions is inconsistent when, in fact, it is consistent. Consider this pair of propositions:

There is an ace or else there is not both a king and a queen.
There is an ace and there is not a king.

As we showed earlier, they should yield an illusion of consistency. A control problem pairs the same disjunctive proposition with a different conjunction:

There is not a king and there is not a queen.

They should yield the correct response that they are consistent. In a similar way, the program predicts illusions of inconsistency and correct responses to their control problems. We tested the occurrence of such illusions in a large sample of participants (489 applicants to a select Italian university). The results showed that the participants were much more accurate with the control problems (83% correct) than with the illusions (27% correct).

Did the participants really construct mental models of the sort proposed by the theory, and did they really interpret the disjunctions as exclusive? A further experiment corroborated both these predictions (Legrenzi et al., 2003). The participants were given a series of problems in which they had to describe a possibility satisfying two propositions or to state that the task was impossible. We used four sorts of problems: illusions of consistency and their controls, and illusions of inconsistency and their controls. If the participants' descriptions deviated systematically from the mental models of the propositions, then they would have refuted the theory. The propositions in the problems referred to various properties, such as "elegant," and to their implicit negations, such as "inelegant." Here is a typical problem:

Only one of the following propositions is true:
The tray is heavy or elegant, or both.
The tray is elegant and portable.
The following proposition is definitely true:
The tray is elegant and portable.
Write a description of the tray

---

where the rubric that only one of two propositions is true amounts to an exclusive disjunction of the two propositions. The program implementing the model theory predicts that reasoners should find the following mental model of the disjunction of the first two propositions:

elegant portable

The conjunction that is definitely true corresponds to this model. Hence, the participants should use this model to formulate a description. Because "heavy" is missing from the model, they should tend to describe the tray with its implicit negation: light, elegant, portable. At its advanced level of expertise, the program constructs each of the following fully explicit models of the disjunction of the first two propositions:

¬ heavy	elegant	¬ portable
heavy	elegant	¬ portable
heavy	¬ elegant	portable
heavy	¬ elegant	¬ portable

The conjunction that is definitely true is inconsistent with each of them, and so it is impossible to frame a description satisfying the propositions. The problem should therefore yield an illusion of consistency. Its control problem pairs the same initial disjunction with a different conjunction that is definitely true: The tray is heavy and elegant, and the participants should respond with the description heavy, elegant, nonportable ("importabile" in Italian). As the fully explicit models show, this response is correct. Analogous conjunctions should yield an illusion of inconsistency and its control. In the experiment, the participants succumbed to the illusions but performed well with control problems. They produced descriptions for the illusions of consistency and their controls that corresponded to the mental models of the propositions. The descriptions for the illusions of consistency, of course, are of nonexistent possibilities; that is, the mental models do not correspond to any of the fully explicit models.

These experiments all support the principle of modeling consistency and its computer implementation. Reasoners appear to evaluate consistency by trying to envisage a possibility compatible with the propositions. If they construct such a model, they can use it to describe an entity consistent with the propositions; if they cannot construct such a model, then they declare that there is no description consistent with the propositions.

The Process of Revision

Previous accounts of how individuals revise their beliefs in the face of inconsistency have proposed that a major factor is the initial credibility of the various beliefs (e.g., Revlin et al., 2001). Fuhrmann (1997) wrote, ". . . when it comes to choosing between candidates for removal, the least entrenched ought to be given up" (p. 24). Harman (1986) distinguishes two main accounts of the entrenchment of beliefs. On the one hand, some theories take into account the "foundations" for a belief, that is, the reasons that support it. If a belief is well founded, then it is unlikely to be lightly abandoned. Various systems of "truth maintenance" in artificial intelligence implement this idea; that is, they aim to keep track of the propositions that support a belief (e.g., de Kleer, 1986; Doyle, 1979). On the other hand, there are systems that take into account only the "coherence" of a belief with other beliefs (see Alchourrón, Gärdenfors, & Makinson, 1985; Gärdenfors, 1988, 1990).

The entrenchment of a belief seems likely to depend on both its coherence with other beliefs and its foundation in terms of reasons that justify it. Thagard (1989, 1992, 2000) has implemented a system that consists of a connectionist algorithm using local representations that assesses alternative hypotheses—for example, about why dinosaurs became extinct—by computing their coherence with the evidence. The user of the system sets up a network of nodes representing the propositions in the competing hypotheses, the propositions in the relevant evidence, and the coherence or

incoherence (with varying degrees of strength) between each pair of propositions. In a process of constraint satisfaction, the computer program then rejects nodes to increase the overall coherence of the system. When the system stabilizes, as it usually does, it shows which hypothesis is more coherent with the evidence. Thagard (2000) reports that human reasoners use less than optimal methods for reaching coherence, but that there are correlations between his system and human performance. Propositions that describe the evidence have a degree of acceptability on their own, which adds a foundational element to the system, but they can be overruled by more coherent propositions.

Psychological evidence corroborates the role of entrenchment in the revision of beliefs. Revlis et al. (1971) gave their participants problems such as

All vertebrates have a backbone.
This amoeba does not have a backbone.

and a proposition that the participants are told is a fact:

This amoeba is a vertebrate.

Given such an inconsistency, the participants tended to believe the general principle, which is common knowledge, rather than the specific proposition, which is not. Other investigators have recently reported similar effects (e.g., Dieussaert et al., 2000; Politzer & Carles, 2001).

The model theory postulates such effects, but it also takes into account the nature of the conflict. A long-standing hypothesis, as we mentioned earlier, is that when a fact is inconsistent with your beliefs, you should make a minimal change to your beliefs to accommodate the fact (see, e.g., Harman, 1986). If so, it is not possible to predict which beliefs you will abandon until you know the conflicting fact. Empirical studies have also corroborated this view. Elio and Pelletier (1997) demonstrated that given an inconsistency between a set of statements and a fact, participants were more likely to abandon a conditional statement than a categorical one. But this effect depended on the nature of the conflict: The participants were more likely to abandon the conditional given a conflict with a modus ponens inference than given a conflict with a modus tollens inference. The investigators suggest that the difference might be a consequence of syntax. A fact that conflicts with a modus ponens inference has the form *not-Q*, and so it also conflicts with the consequent clause of the conditional *If P then Q*. A fact that conflicts with a modus tollens inference has the form *P*, which does not conflict with the antecedent clause of the conditional (see also Revlin et al., 2001). In neither case is there a real contradiction. The syntactic conflict, however, may lead individuals to reject the conditional in the modus ponens case.

What distinguishes the model theory's mismatch principle from this alternative account is that it predicts that individuals should retract the proposition that has mental models conflicting with the fact or failing to represent the fact, and that these effects depend on the level of expertise at which individuals represent propositions. In a series of studies, Hasson and Johnson-Laird (2003) examined the relative believability of conditional and categorical propositions in the form of different speakers' assertions of the premises for modus ponens and modus tollens inferences. The participants' task was to rate the relative believability of the two propositions on a single scale, in which 1 corresponded to complete belief in one speaker's assertion and 6 corresponded to complete belief in the other speaker's assertion. When the two propositions occurred without any conflicting fact, individuals tended to reject the conditional proposition. For instance, they reported that the conditional

If the professor is correct, then the gene releases a signal when lactose enters the cell.

seems hypothetical and open to doubt. But when the two assertions occurred with a fact that conflicted with their consequence, individuals tended to reject the categorical proposition. This result shows that entrenchment alone cannot predict the revision of beliefs in the face of inconsistency. Results also corroborate the mismatch principle. As Elio and Pelletier (1997) observed, individuals rejected the conditional when the fact conflicted with its explicit mental model, and so the conditional is less believable in a conflict with a modus ponens inference than in a conflict with a modus tollens inference. This effect is modulated both by negation and by the strategies that individuals use to cope with inconsistencies (Hasson & Johnson-Laird, 2003).

In a recent study, Byrne and Walsh (2002) obtained results that appear to be contrary to the mismatch principle. Their participants were more inclined to believe conditionals when the facts were inconsistent with modus ponens than when they were inconsistent with modus tollens. They used a different experimental procedure, however. Their participants had to draw an explicit conclusion before they were shown the inconsistent fact. This procedure should bias the participants to use the strategy in which they detect the inconsistency between the conclusion of the two statements and the fact. In an unpublished study, Hasson and Walsh (2003) compared the two procedures in a single experiment. The experiment replicated both sets of results. Different procedures elicit different reasoning strategies, which in turn affect the relative believability of statements.

Could the effects of mismatch be purely syntactic, as Elio and Pelletier (1997) suggested? A study of more complex inconsistencies examined this hypothesis (Girotto, Johnson-Laird, Legrenzi, & Sonino, 2000). A preliminary experiment corroborated the mismatch principle with problems such as

Paolo says, "The President owns a villa and a swimming pool, or else he owns a plane."
Vittorio says, "The President owns a villa and a swimming pool."
But you happen to know that the President owns a plane.
According to you, who asserted a false proposition?

The results corroborated the mismatch principle, which predicts that individuals should reject Vittorio's assertion (98 out of the 111 participants conformed to it more often than not, and there were 7 ties). In other cases, as the mismatch principle predicts, individuals tended to reject a conjunction rather than a more complex proposition. Further experiments have shown that reasoners indeed rely on mental models rather than syntax. Consider, for example, the following problem:

Paolo says, "If the President owns a villa and a swimming pool, then he owns either a plane or else a Ferrari."

Vittorio says, "The President owns a villa and a swimming pool."

But you happen to know that he owns a plane and a Ferrari.

As the program shows, the participants should judge correctly that the propositions are inconsistent, and it predicts that they should retract the complex proposition. But they cannot do so merely on the basis of a surface mismatch of owning a plane and a Ferrari. They have to grasp that the conjunction is inconsistent with the exclusive disjunction in the consequent of the conditional. A conjunction is consistent with many other connectives, such as a conditional, and to appreciate its inconsistency with an exclusive disjunction, reasoners have to build models. The results of a variety of problems corroborated the mismatch principle.

The Generation of Explanations

So far, we have assessed the model theory and other accounts of how individuals detect inconsistencies and revise propositions in order to achieve consistency. We now turn to the generation of diagnostic explanations to resolve inconsistencies. Researchers have examined how people diagnose faults (e.g., Rasmussen, 1981; Rouse & Hunt, 1984) and illnesses (e.g., Kuipers & Kassirer, 1984; Patel, Groen, & Arocha, 1990). Both tasks are similar to the explanations of inconsistencies. Not surprisingly, however, these studies have not addressed the predictions of the model theory, and so we have carried out a series of studies designed to test its main predictions: Individuals should be biased in favor of causal explanations; they should understand these relations to be deterministic rather than probabilistic; they should show the predicted preference for cause-and-effect explanations over minimal explanations of cause alone or effect alone; and the mismatch principle should influence their preferences.

If knowledge generally takes precedence over inconsistent propositions, then the explanatory mechanism, which depends on knowledge, should dominate the ability to make deductions, which depends on propositions. Given a choice between an abduction and a deduction, individuals should tend to make the abduction. It is not easy to test this prediction in an experiment free from confounds, but we were able to do so in a study that was also designed to examine the sorts of explanations that individuals spontaneously create. We presented the participants with problems that they could solve either by making a logical deduction or by generating a causal explanation. One problem, for instance, was

If a pilot falls from a plane without a parachute then the pilot dies.

This pilot did not die.

Why not?

Like the other problems, this problem has the form:

If A then B. Not B. Why not?

The participants could answer the question by deducing a modus tollens conclusion (*not A*):

This pilot did *not* fall from a plane without a parachute.

Alternatively, they could assume that the pilot fell from the plane without a parachute—a Gricean implicature perhaps—and offer a causal explanation for the pilot's survival. Overwhelmingly, they answered with causal explanations (77%) as opposed to deductions or other responses (23%), and no participant went against this trend. In the case of the pilot, for example, they generated the following sorts of explanations:

The plane was on the ground and so the pilot fell only a short distance.

The pilot fell into a deep snowdrift and so his (sic) fall was cushioned.

The fact that they chose explanations rather than deductions is presumably because the Gricean implicature is salient, and the creation of an explanation is easier than a modus tollens deduction (see, e.g., Evans et al., 1993). The deduction, however, calls for no change in beliefs, whereas the abduction calls for accepting at least one new belief.

Separate studies[2] confirmed that individuals tend to think of a series of explanations of the same inconsistency in correlated orders—a tendency that presumably reflects the knowledge available to members of the same culture. And they were able to explain random conjunctions of events selected from separate stories. Hence, causal explanations cannot merely be retrieved from memory. They must be inferred from separate components of causal knowledge. But what are these components?

According to the model theory, causal relations refer to sets of temporally ordered possibilities (see the previous section of the article). This account has been corroborated in a series of studies (see Goldvarg & Johnson-Laird, 2001). In several of these experiments, the participants had to list what was possible and what was impossible for different propositions, including those stating causes and those stating enabling conditions. As the theory predicts, they tended to list either the three possibilities for an ordinary cause or the two possibilities for a unique cause; and they tended to list either the four possibilities for an ordinary enabling condition or the three possibilities for a unique enabling condition.

A further study showed that individuals can distinguish causes from enabling conditions when they occur in the same scenario (Goldvarg & Johnson-Laird, 2001). For example, with the following scenario:

Given that there is good sunlight, if a certain new fertilizer is used on poor flowers, then they grow remarkably well. However, if there is not good sunlight, poor flowers do not grow well even if the fertilizer is used on them.

Most participants judged that the sunlight was the enabling condition and the fertilizer was the cause of the flowers' growth. But, given the following scenario:

Given the use of a certain new fertilizer on poor flowers, if there is good sunlight, then the flowers grow remarkably well. However,

[2] These studies were carried out in collaboration with Tony Anderson.

if the new fertilizer is not used on poor flowers, they do not grow well even if there is good sunlight.

Most participants judged that the fertilizer was the enabling condition and the sunshine was the cause of the flowers' growth. They made these judgments even when the order of the clauses was changed so that the first entity referred to was the cause instead of the enabling condition. In these scenarios, cause and enabling conditions differ in meaning, and naive individuals are able to identify them reliably. This phenomenon corroborates the model theory, but it counts against the need for a mechanism in assigning causal roles in a scenario. Most people do not know the mechanism underlying the growth of plants. And whatever it is, it can hardly underlie the interpretations of both scenarios.

A further set of studies showed the predicted differences in the inferences that individuals made from causal premises and from enabling premises. For example, given a problem such as:

Eating protein will cause her to gain weight.
She will eat protein.
Will she gain weight?

Most participants inferred that she will gain weight. They refrained from this inference, however, when the first premise stated an enabling condition:

Eating protein will allow her to gain weight.

The results in general are contrary to claims that causes and enabling conditions do not differ in meaning or logic. They are also contrary to probabilistic theories of causation. On a probabilistic analysis of *A will cause B,* no event should be judged "impossible," and no definite conclusion should follow logically from the further categorical premise *A.* For the case against other theories of causation, see Goldvarg and Johnson-Laird (2001).

As we showed earlier, the model theory predicts that a chain consisting of a cause and an effect should be a better resolution of an inconsistency than the cause alone, which in turn should be better than the effect alone. In addition, the mismatch principle predicts that explanations of a conflict with modus ponens should tend to eliminate the conditional premise rather than the categorical premise. The prediction that inferences from causes to effects should seem more plausible than inferences from effects to causes fits the results of previous studies. As Tversky and Kahneman (1982) established, conditional propositions in which the antecedent is the cause of a consequent effect, such as: "A girl has blue eyes if her mother has blue eyes," are judged as more probable than conditional propositions in which the antecedent is evidence for the cause stated in the consequent: "The mother has blue eyes if her daughter has blue eyes." We also carried out direct tests of the principle of causal knowledge.

In a preliminary study, we gave the participants a series of 20 problems from a variety of domains, in which there was an inconsistency with a modus ponens inference, for example:

If a person pulls the trigger, then the gun will fire. Someone has pulled the trigger, but the gun did not fire. Why not?

The participants had to give a single explanation in their own words of why each consequent had not occurred. Each problem elicited a variety of different explanations, with a mean of 4.75 different explanations per problem. As the mismatch principle predicts, however, the vast majority of explanations amounted to retractions of the conditional proposition (90% of trials) rather than the categorical proposition. Only on 2% of trials were the participants unable to come up with an explanation (the *Marie Celeste* phenomenon).

In a second experiment, the participants had to rank order the probabilities of putative explanations of the inconsistencies, which were based on those from the previous study. As the model theory predicts, the participants showed an overwhelming and highly significant tendency to rank order the probabilities of the sets of putative explanation in the following order, starting with the most probable explanation:

1. Cause and effect: A prudent person had unloaded the gun and there were no bullets in the chamber.
2. Cause alone: A prudent person had unloaded the gun.
3. Effect alone: There were no bullets in the chamber.
4. Rejection of the categorical proposition: The person didn't really pull the trigger.
5. Noncausal conjunction: The gun was heavy, and there were no bullets in the chamber.

The cause and effect is a conjunction, so the noncausal conjunction was included as a control. One weakness of the experiment was that all the participants received the 20 scenarios in the same order. However, we replicated the results in a subsequent experiment in which they were presented in four different orders. The rankings are instances of the "conjunction" fallacy (Tversky & Kahneman, 1983), because the cause-and-effect conjunctions were ranked as more probable than their individual constituent propositions. The results also showed that individuals are unlikely to accommodate a new fact with an invariable minimal change to their existing beliefs. The acceptance of a conjunction calls for a greater change than the acceptance of just one of its constituent propositions. The best explanation is not always a minimal one.

To what extent is the trend in the previous studies a reflection of the attraction of causal explanations as opposed to the attraction of explanations that rule out the conditional proposition? To answer this question, we carried out an experiment in which the causal explanation ruled out, not the conditional, but the categorical proposition. For example, in the gun example, the explanations concerned why the person had not pulled the trigger. In this case, the predicted trend occurred over the following sorts of explanation:

1. Cause and effect: The person was semiparalyzed, and he was not able to move his fingers with sufficient strength.
2. Cause alone: The person was semiparalyzed.
3. Effect alone: The person was unable to move his fingers with sufficient strength.

But, in this study, an explanation that retracted the conditional was ranked in probability just after the cause and effect.

4. Retraction of the conditional: There were no bullets in the chamber.

The participants ranked as least probable the control conjunction

5. Noncausal conjunction: Paolo was proud and he was not able to move his fingers with sufficient strength.

Hence, explanations that rule out the conditional proposition remain attractive even in the context of competing explanations that rule out the categorical proposition.

The mismatch principle predicts, as we showed earlier, that individuals should be more likely to reject a categorical proposition when an ordinary conditional is replaced by a biconditional, such as:

If and only if a person pulled the trigger then the gun fired.

The biconditional is more likely to elicit models of the following possibilities:

pulled-trigger	fired
¬ pulled-trigger	¬ fired.

The fact that the gun did not fire matches the second model, and so reasoners should be more inclined to accept an explanation that retracts the categorical proposition. We have carried out three experiments that show a concomitant shift in the rank order of putative explanations. One experiment made a direct comparison between indicative conditionals and biconditionals, counterbalancing their contents from one participant to another. Every participant ranked the retraction of the categorical as having a higher probability in the scenarios based on the biconditionals than in the scenarios based on the conditionals.

In sum, the plausibility of an explanation appears to reflect two tendencies. First, following the principle of causal knowledge, individuals find causal chains that resolve inconsistencies highly plausible even though they are not minimal. Second, following the mismatch principle, they find explanations that rule out the conditional more plausible than explanations that rule out the categorical proposition unless they flesh out their models of the conditional explicitly.

Discussion

Life confronts you with surprises. They clash with the consequences of your beliefs. When you detect an inconsistency, you try to reason to consistency. You give up conclusions in the face of facts to the contrary, seek to modify sets of propositions to make them consistent, and attempt to create a plausible diagnosis of what has gone wrong. In contrast, logical deductions are cumulative and monotonic: As you discover more premises, so you can deduce more conclusions. Are there any deductions in real life that are indefeasible, that never stand open to correction? In fact, no deduction based on contingent propositions—as opposed to necessary truths or axiomatic assumptions—can be guaranteed forever. Any conclusion of a valid inference may be overturned by decisive evidence to the contrary. Skeptics might take this fact to minimize the importance of deduction in daily life. The opposite moral should be drawn. Only conflicts between facts and the valid consequences of your beliefs should force you to reason to consistency. Human reasoning, however, is often carried out within a protected mental environment—a kind of intellectual laboratory—in which conclusions can be assessed in relation to beliefs, knowledge, and facts. This process of evaluation can lead to various outcomes, including the acceptance of a conclusion or its retraction. Students of artificial intelligence have developed many computer programs for nonmonotonic reasoning in which prior conclusions are abandoned in the light of later information, and for the revision of beliefs in the face of contradiction. Only a handful of cognitive psychologists have studied these processes (e.g., Dieussaert et al., 2000; Elio & Pelletier, 1997; Politzer & Carles, 2001; Revlis & Hayes, 1972).

Oaksford and Chater (1991) have argued that reasoning to consistency should be a computationally tractable process. We are sympathetic to this proposal, but there are grounds for caution. A function is tractable if a correct result can be computed for any possible input, and the time and memory requirements for the computation increase only in proportion to some polynomial of the size of the input. The operations of the human parser for natural language appear to be tractable, because people do not lag ever further behind in understanding longer and longer sentences. In contrast, reasoning with negation and sentential connectives, such as *if* and *or*, is intractable, and human reasoning of this sort soon breaks down with inferences based on an increasing number of propositions. It also breaks down in the detection of inconsistencies. In our experiments, the participants generated a variety of explanations. Sometimes, these explanations emerged rapidly and were highly plausible, though no clear criterion exists for what counts as a good putative explanation. A still more serious obstacle to claims of tractability is the occasional failure to come up with any explanation—a case of the *Marie Celeste* phenomenon. This phenomenon shows that reasoning to consistency does not always yield an outcome, let alone a plausible one. Rapid and effortless inferences in certain cases tell one nothing about tractability overall. In sum, it is doubtful whether individuals use a tractable algorithm for reasoning to consistency. Their algorithm can work well with problems on a human scale, but it will be overwhelmed by complex problems.

Our goal has been to advance a theory of how people reason from inconsistency to consistency. This account depends on existing principles of the model theory. In particular, it depends on the principle of truth: Individuals represent situations by constructing sets of mental models in which each model represents what is true in a possibility. The existing theory also allows that when individuals construct mental models, they can make arbitrary assumptions or assumptions by default, which they can revise, if necessary, in the light of later information (Johnson-Laird & Byrne, 1991). These assumptions account in principle for reasoning to consistency when an inconsistency arises as a result of an arbitrary or default assumption (Brewka et al. 1997). But, inconsistencies arise in many other circumstances. In our main example, you believe that if Paolo has gone to get the car, then he will be back in five minutes, and that Paolo has gone to get the car. But he fails to return in five minutes. In this case, the premises are hardly default assumptions. It is necessary instead to reason to consistency. We distinguished three main processes in such reasoning. First, you

must detect the inconsistency. Paolo's failure to return conflicts with a logical consequence of your beliefs. Other inconsistencies are subtler, and unless you detect them, you may blithely continue to hold erroneous beliefs. Second, you must revise your beliefs. Perhaps your categorical assumption that Paolo has gone to get the car is false. Third, you must try to create a diagnosis that resolves the inconsistency. You envisage various scenarios about what has happened. In the actual historical case, he had trouble starting the car.

The present theory adds three new principles to explain reasoning to consistency. To account for the detection of inconsistencies, the theory postulates

> *Modeling consistency:* People evaluate consistency on the basis of mental models of the relevant propositions. If they find a model that satisfies all the propositions, they judge the set to be consistent; otherwise, they judge it to be inconsistent.

We have implemented the theory in a computer program that reasons to consistency. It was the program which revealed that the principle of truth predicts the occurrence of illusions of consistency and inconsistency. The evidence, summarized in the previous section, corroborated the principle (Johnson-Laird et al., 2000).

Previous theories of belief revision have often emphasized the notion that individuals abandon the least credible proposition in the case of an inconsistency (e.g., Revlis & Hayes, 1972). The model theory allows that entrenchment of beliefs is important, but it also emphasizes the role of reasoning in the revision of beliefs (Hasson & Johnson-Laird, 2003). It postulates

> *Mismatches:* When an inconsistency occurs between facts and a set of plausible propositions taken collectively, individuals tend to retract a proposition that has mental models that conflict with the facts or that otherwise fails to represent the facts.

In many cases, the appearance is deceptive: Fully explicit models show that the proposition is consistent with the facts. Even when individuals realize that there is no genuine inconsistency, they are nevertheless disposed to retract the mismatching proposition. Estimates of the probability of a proposition can accordingly switch from a situation in which they are presented alone to a situation in which they are presented with a conflicting fact (Hasson & Johnson-Laird, 2003). When a fact conflicts with a set of propositions, which contains no highly dubious proposition, individuals tend to retract whichever unique proposition has only mental models that conflict with the fact or that fail to represent the fact. Such mismatches might merely concern surface clauses in sentences. However, as the evidence in the previous section shows, individuals assess mismatches between models rather than between surface clauses.

To explain how individuals create diagnoses to resolve inconsistencies, the theory assumes that individuals—at least in Western culture and possibly all cultures—search for causal accounts that explain the origins of the inconsistency. The theory accordingly postulates

> *Causal knowledge:* Different models of temporally ordered possibilities underlie knowledge of causes and enabling conditions. These possibilities can be used to construct causal chains. The optimal chain consists of a cause and its effect. In the resolution of an inconsistency, such a chain takes precedence over the models of the propositions and explains the inconsistent fact.

Our studies corroborated the model theory's account, showing that individuals distinguish between the meanings and logical consequences of *A will cause B* and *A will allow B* (e.g., Goldvarg & Johnson-Laird, 2001). Individuals tend to generate causal explanations rather than to make deductions to resolve inconsistencies. Contrary to the common philosophical doctrine, going back at least to James (1907), individuals do not always make minimal changes to their beliefs. They judge as more probable an explanation of an inconsistency that describes both a cause and an effect. An explanation consisting of the effect alone is ad hoc. For instance, one explanation of Paolo's failure to return in five minutes might be that it took him a long time to drive back from the parking garage. The explanation is unmotivated. It, in turn, stands in need of explanation. Hence, a more plausible explanation would be that he ran into a complicated one-way system and so it took him a long time to drive back from the car park. Our experiments corroborated the theory: The participants rated the conjunction of a cause and an effect as more probable than the cause alone, which they rated as more probable than the effect alone (see the previous section). This trend shows that individuals do not invariably prefer explanations that yield minimal changes to their beliefs, as does any instance of the conjunction fallacy in explanations (Tversky & Kahneman, 1983).

The principle of causal explanation interacts with the mismatch principle. When the propositions contain a biconditional, such as

> If and only if someone pulled the trigger, then the gun fired

in place of an ordinary conditional, then the ratings of the putative explanations alter. Individuals are more likely to give a higher probability to an explanation that rules out the categorical proposition; that is, they tend to accept the biconditional. The phenomenon is predicted by the mismatch principle, because individuals are more likely to flesh out their models of a biconditional to include the possibility that matches the fact (Johnson-Laird & Byrne, 1991).

Which elements in the model theory of reasoning to consistency are robust, and which elements are liable to change (nonmonotonically) as a result of future research? Individuals can detect inconsistencies in a self-consciously critical frame of mind, as participants in experiments or, say, as reviewers of papers. In this frame of mind, they can search for conflicts between one proposition and another, and among a set of propositions as a whole. But, in a more relaxed setting in daily life, they can also notice conflicts between the consequences of their beliefs and what actually happens. In both situations, it seems plausible that they are envisaging possibilities; that is, they are constructing mental models. In an experimental setting, the occurrence of illusions is excellent evidence in support of modeling consistency. Illusions in deduction occur in daily life, and we suspect that their counterparts in consistency also occur. We are all likely to succumb to illusions

because their detection is difficult (and computationally intractable).

The mismatch principle seems less secure. Our evidence in support of the principle is good, but why should people doubt a proposition because it has mental models conflicting with a fact or failing to represent it? In one construal of the principle, such retractions are also akin to illusions. When a proposition has only mental models conflicting with the fact, it seems inconsistent with the fact, even if it is not. Its compatibility, however, depends on a possibility that is not represented in a mental model, but only in a fully explicit model. Individuals usually overlook these models. Likewise, when they retract a proposition because it fails to represent the fact, it is important to remember that the other propositions in the set must match the fact. That is, these propositions have mental models that represent the fact, and so they seem to be compatible with it.

The mismatch principle seems plausible as an account of laboratory performance, but its role in daily life is less certain. We have couched the principle in broad terms: Individuals either accept or reject beliefs. But, of course, they tend to hold beliefs with varying degrees of confidence, and so a more refined mismatch principle might predict modifications in the strength of beliefs. Psychologists have increasingly studied the effects of uncertainty in the premises on estimates of the likelihood of conclusions (see, e.g., Stevenson & Over, 1995, 2001), and even formulated accounts of reasoning that are intrinsically probabilistic (Oaksford, Chater, & Larkin, 2000). A plausible conjecture is accordingly that in the face of inconsistency individuals do not merely accept or retract their beliefs but rather modify the strength of their beliefs (Politzer & Carles, 2001). Various potential measures of strength of belief are feasible, but the most natural one is probability. The model theory can account for the probability of a belief (Johnson-Laird et al., 1999). It depends on the proportion of equiprobable models in a partition in which the belief holds, or alternatively, on numerical probabilities attached to these models. Depending on the context, individuals may use either sort of representation. A more refined version of the mismatch principle can therefore be formulated in terms of probabilities. But, if beliefs are merely probable, then evidence is much less likely to be inconsistent with them. Suppose that you infer that Paolo should return in five minutes, and your degree of belief is equivalent to a probability of .75. If Paolo fails to return in five minutes, then there is no direct conflict with your belief, which has a probability of .25 of being false. The introduction of probabilities accordingly raises a new problem for a theory of reasoning to consistency. What probability or degree of confidence do you have to have in a proposition, *A*, for the facts, *not-A*, to count as inconsistent?

Another concern about mismatches in daily life is that, as we have argued, individuals normally worry about how to explain an inconsistency, and allow the explanation to have as a side effect the revision of beliefs. In a preliminary study, we solicited the participants' spontaneous explanations of conflicts. The majority of their explanations amounted to retractions of the conditional premises in a series of 20 modus ponens arguments. This finding is exactly what the mismatch principle predicts. Yet, it takes work to determine which premise an explanation rules out. Could the result have occurred for reasons other than mismatch? Intuitive judgments of probability might have come into play in this case: Conditionals may seem less probable than categorical propositions. But, other studies corroborate the mismatch principle, and in the appropriate conditions, individuals rate conditionals as more believable than categoricals (e.g., Hasson & Johnson-Laird, 2003).

The premium our culture places on causal explanation is indubitable. It is the implications of causal knowledge that are more surprising. They yield a preference for a causal chain to resolve an inconsistency. This preference is contrary to the seemingly plausible assumption that individuals should make minimal changes to their beliefs in order to accommodate inconsistencies. To accept both a cause and an effect as new beliefs is plainly not a minimal step. Yet, the model theory predicts the preference on the grounds that an effect without a cause lacks a motivation, and so it seems ad hoc. Individuals prefer a causal account of an inconsistency with modus ponens that rules out the conditional proposition, and they downplay as improbable retractions of the categorical proposition. Change the conditional to a biconditional, however, and an explanation that rules out the categorical becomes much more plausible, again, another corroboration of the mismatch principle, and of the notion that knowledge modulates the interpretation of propositions.

The conclusion is clear. The experimental results support the model theory of reasoning to consistency. Reasoners detect inconsistencies in their mental models of facts and their beliefs. They retract whatever mismatches the fact. Direct clashes and singularly dubious propositions aside, they retract whichever proposition has only mental models conflicting with the fact or otherwise failing to represent the fact. They use their available knowledge—in the form of causal models—to try to create a causal scenario that makes sense of the facts of the matter. Their reasoning may resolve the inconsistency. It may yield an erroneous model of the situation. It may yield no model at all.

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic, 50,* 510–530.

Bara, B., Bucciarelli, M., & Johnson-Laird, P. N. (1995). The development of syllogistic reasoning. *American Journal of Psychology, 108,* 157–193.

Barres, P., & Johnson-Laird, P. N. (2003). On imagining what is true (and what is false). *Thinking & Reasoning, 9,* 1–42.

Barwise, J. (1993). Everyday reasoning and logical inference. *Behavioral and Brain Sciences, 16,* 337–338.

Bauer, M. I., & Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychological Science, 4,* 372–378.

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic*. Mahwah, NJ: Erlbaum.

Brewka, G., Dix, J., & Konolige, K. (1997). *Nonmonotonic reasoning: An overview*. Stanford, CA: CSLI Publications.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning. *Cognitive Science, 23,* 247–303.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575.

Byrne, R. M. J., & Walsh, C. R. (2002). Contradictions and counterfactuals: Generating belief revisions in conditional inference. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 160–165). Mahwah, NJ: Erlbaum.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104,* 367–405.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545–567.

Cheng, P. W., & Novick, L. R. (1991). Causes versus enabling conditions. *Cognition, 40,* 83–120.

Cook, S. A. (1971). The complexity of theorem proving procedures. *Proceedings of the Third Annual Association of Computing Machinery Symposium on the Theory of Computing, 3,* 151–158.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition, 19,* 274–282.

de Kleer, J. (1986). An assumption-based TMS. *Artificial Intelligence, 28,* 127–162.

Dieussaert, K., Schaeken, W., De Neys, W., & d'Ydewalle, G. (2000). Initial belief state as a predictor of belief revision. *Current Psychology of Cognition, 19,* 277–288.

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence, 12,* 231–272.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3–19.

Elio R. (1997). What to believe when inferences are contradicted. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 211–216). Hillsdale, NJ: Erlbaum.

Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science, 21,* 419–460.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T., & Over, D. (1996). *Rationality and reasoning*. Hove, England: Psychology Press.

Fuhrmann, A. (1997). *An essay on contraction*. Stanford, CA: CSLI Publications.

Gärdenfors, P. (1988). *Knowledge in flux.* Cambridge, MA: MIT Press.

Gärdenfors, P. (1990). The dynamics of belief systems: Foundations vs. coherence theories. *Revue Internationale de Philosophie, 172,* 24–46.

Garnham, A., & Oakhill, J. V. (1994). *Thinking and reasoning*. Oxford, England: Basil Blackwell.

Ginsberg, M. L. (1987). Introduction. In M. L. Ginsberg (Ed.), *Readings in nonmonotonic reasoning* (pp. 1–23). Los Altos, CA: Morgan Kaufmann.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition, 78,* 247–276.

Girotto, V., Johnson-Laird, P. N., Legrenzi, P., & Sonino, M. (2000). Reasoning to consistency: How people resolve logical inconsistencies. In J. Garcia-Madruga, M. Carriedo, & M. J. Gonzalez-Labra (Eds.), *Mental models in reasoning* (pp. 83–97). Madrid, Spain: UNED.

Goldvarg, Y., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science, 25,* 565–610.

Greene, B. (2000). *The elegant universe*. New York: Vintage Books.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: Bradford Book.

Hart, H. L. A., & Honoré, A. M. (1985). *Causation in the law* (2nd ed.). Oxford, England: Clarendon Press.

Hasson, U., & Johnson-Laird, P. N. (2003). *How reasoning changes your beliefs*. Manuscript submitted for publication.

Hasson, U., & Walsh, C. (2003). [The resolution of MP and MT inconsistencies]. Unpublished study, Princeton University.

Hilton, D. J., & Erb, H.-P. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking & Reasoning, 2,* 273–308.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1–55.

James, W. (1907). *Pragmatism—A new name for some old ways of thinking*. New York: Longmans, Green & Co.

Johnson-Laird, P. N. (1993). *Human and machine thinking*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646–678.

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory & Cognition, 31,* 1105–1113.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M. S. (2000, April 21). Illusions in reasoning about consistency. *Science, 288,* 531–532.

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., Legrenzi, M. S., & Caverni, J.-P. (1999). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62–88.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternative. *Psychological Review, 93,* 75–88.

Kuipers, B., & Kassirer, J. P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8,* 363–385.

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition, 49,* 37–66.

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (2003). Models of consistency. *Psychological Science, 14,* 131–137.

Lepper, M. R., Ross, L., & Lau, R. R. (1986). Persistence of inaccurate beliefs about the self: Perseverance effects in the classroom. *Journal of Personality and Social Psychology, 50,* 482–491.

Levi, I. (1991). *The fixation of belief and its undoing: Changing beliefs through inquiry*. New York: Cambridge University Press.

Markovits, H. (1984). Awareness of the "possible" as a mediator of formal thinking in conditional reasoning problems. *British Journal of Psychology, 75,* 367–376.

Mill, J. S. (1874). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific evidence* (8th ed.). New York: Harper. (Original work published 1843)

Morris, M. W., Nisbett, R. E., & Peng, K. (1995). Causal attribution across domains and cultures. In C. Lewis, D. Premack, & D. Sperber (Eds.), *Causal cognition* (pp. 577–614). New York: Oxford University Press.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6,* 1–38.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 883–899.

Ormerod, T. C., Manktelow, K. I., & Jones, G. V. (1993). Reasoning with three types of conditional: Biases and mental models. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46*(A), 653–678.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition, 18,* 394–406.

Peirce, C. S. (1955). Abduction and induction. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 150–156). New York: Dover. (Original work published 1903)

Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce* (C. Hartshorne, P. Weiss, & A. Burks, Eds.). Cambridge, MA: Harvard University Press.

Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking & Reasoning, 7,* 217–234.

Polk, T. A., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533–566.

Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 241–258). New York: Plenum Press.

Rehder, B., & Hastie, R. (1996). The moderating influence and variability on belief revision. *Psychonomic Bulletin & Review, 3,* 499–503.

Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence, 13,* 81–132.

Revlin, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rending. *Memory & Cognition, 29,* 1196–1208.

Revlis, R. (1974). Prevarication: Reasoning from false assumptions. *Memory & Cognition, 2,* 87–95.

Revlis, R., & Hayes, J. R. (1972). The primacy of generalities in hypothetical reasoning. *Cognitive Psychology, 3,* 268–290.

Revlis, R., Lipkin, S. G., & Hayes, J. R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning and Verbal Behavior, 10,* 86–91.

Rips, L. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. A. Shweder (Ed.), *Fallible judgement in behavioral research*: *New directions for methodology of social and behavioral science* (Vol. 4, pp. 17–36). San Francisco, CA: Jossey-Bass.

Rouse, W. B., & Hunt, R. M. (1984). Human problem solving in fault diagnosis tasks. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (pp. 195–222). Greenwich, CT: JAI Press.

Schaeken, W. S., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition, 60,* 205–234.

Schlottmann, A., & Anderson, N. H. (1995). Belief revision in children: Serial judgment in social cognition and decision-making domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1349–1364.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 613–643.

Stevenson, R. J., & Over, D. E. (2001). Reasoning from uncertain premises: Effects of expertise and conversational context. *Thinking & Reasoning, 7,* 367–390.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.

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

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Turnbull, W., & Slugoski, B. R. (1988). Conversational and linguistic processes in causal attribution. In D. Hilton (Ed.), *Contemporary science and natural explanation*: *Commonsense conceptions of causality* (pp. 66–93). Brighton, England: Harvester Press.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgements under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117–128). Cambridge, England: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 292–315.

Van der Henst, J.-B., Yang, Y., & Johnson-Laird, P. N. (2002). Strategies in sentential reasoning. *Cognitive Science, 26,* 425–468.

Wason, P. C. (1964). The effect of self-contradiction on fallacious reasoning. *Quarterly Journal of Experimental Psychology,16,* 30–34.

Zhang, J., Johnson, T. R., & Wang, H. (1997). The relation between order effects and frequency learning in tactical decision making. *Thinking & Reasoning, 4,* 123–145.

Appendix A

Output of the First Stage of the Computer Program That Uses Mental Models and Fully Explicit Models to Check the Consistency of the Set of Propositions, *Ace ore not comma king and queen; Ace and not king*

1. Mental models:
   The proposition, *Ace ore not comma king and queen,* is satisfied by the mental model
   ACE
   The propositions including *Ace and not king* are satisfied by the mental model
   ACE ¬ KING
   Hence, individuals should judge that the propositions are consistent
   Mental models of the proposition, *Ace ore not comma king and queen:*
   ACE
   KING ¬ QUEEN
   ¬ KING QUEEN
   ¬ KING ¬ QUEEN
   Mental models of the proposition, *Ace and not king:*
   ACE ¬ KING
2. Fully explicit models:
   The proposition, *Ace ore not comma king and queen,* is satisfied by the fully explicit model
   ACE KING QUEEN
   The propositions including *Ace and not king* have no fully explicit model
   Hence, fully explicit models show that the propositions are *not* consistent
   Fully explicit models of the proposition, *Ace ore not comma king and queen:*
   ACE KING QUEEN
   ¬ ACE KING ¬ QUEEN
   ¬ ACE ¬ KING QUEEN
   ¬ ACE ¬ KING ¬ QUEEN
   Fully explicit models of the proposition, *Ace and not king,* have the mental models
   ACE ¬ KING

Note. The program predicts that individuals who use mental models should judge that the set is consistent, whereas, as the fully explicit models show, it is in fact inconsistent. "Ore" denotes an exclusive disjunction, and "comma" indicates that the negation applies to the conjunction as a whole.

(*Appendixes continue*)

Appendix B

Output of the Second Stage of the Program, Which Implements the Mismatch Principle to Determine Which Proposition to Abandon in Light of a Contradiction With the Facts

1. Mental models used to check mismatch:
   Mental models of the propositions
   PLANE-ON-COURSE RADAR-SHOWS-WATER
   . . .
   PLANE-ON-COURSE
   Fact model:
   ¬ RADAR-SHOWS-WATER
   Fact model mismatches models of *If plane-on-course, then radar-shows-water*
   Fact model is not represented in models of *plane-on-course*
   So, reject *If plane-on-course, then radar-shows-water*
   Revised models:
   Facts: PLANE-ON-COURSE ¬ RADAR-SHOWS-WATER
   Counterfactual possibilities: PLANE-ON-COURSE RADAR-SHOWS-WATER
   . . .
2. Fully explicit models used to check mismatch:
   Fully explicit models of the propositions
   PLANE-ON-COURSE RADAR-SHOWS-WATER
   ¬ PLANE-ON-COURSE RADAR-SHOWS-WATER
   ¬ PLANE-ON-COURSE ¬ RADAR-SHOWS-WATER
   PLANE-ON-COURSE
   Fact model:
   ¬ RADAR-SHOWS-WATER
   Fact model matches model of *If plane-on-course, then radar-shows-water*
   Fact model is not represented in model of *plane-on-course*
   So reject proposition *plane-on-course*
   Revised models:
   Facts: ¬ PLANE-ON-COURSE ¬ RADAR-SHOWS-WATER
   Counterfactual possibilities: PLANE-ON-COURSE

Note. The example is a case of a contradiction with an inference in the form of modus ponens: *If plane-on-course, then radar-shows-water. Plane-on-course. Not radar-shows-water.*

Appendix C

Output of the Third Stage of the Computer Program, Which Uses a Representation of General Knowledge to Construct a Causal Chain to Resolve the Inconsistency,
If someone pulled the trigger, then the gun fired.
Someone pulled the trigger. But the gun did not fire.

1. Mental models used to resolve inconsistency:
   Explicandum:
      PULLED-TRIGGER ¬ GUN-FIRED
   Explicans:
      PULLED-TRIGGER GUN-BROKEN ¬ GUN-FIRED
   The explanation of the inconsistency from available knowledge:
      Pulled-trigger, and gun-broken, and so it is not the case gun-fired
   Explicandum:
      GUN-BROKEN
   Explicans:
      GUN-DROPPED GUN-BROKEN
   Why is the following the case: gun-broken?
      Gun-dropped, and so gun-broken
2. Fully explicit models used to resolve inconsistency:
   Explicandum:
      ¬ PULLED-TRIGGER ¬ GUN-FIRED
   Explicans:
      ¬ PULLED-TRIGGER ¬ ENOUGH-STRENGTH ¬ GUN-FIRED
   The explanation of the inconsistency from available knowledge:
      It is not the case that pulled-trigger, and it is not the case that enough-strength, and so it is not the case that gun-fired
   Explicandum:
      ¬ ENOUGH-STRENGTH
   Explicans:
      PARTIAL-PARALYSIS ¬ ENOUGH-STRENGTH
   Why is the following the case: It is not the case that enough-strength?
      Partial-paralysis, and so it is not the case enough-strength

Note. The program follows the mismatch principle; so with mental models, it rejects the conditional, whereas with fully explicit models, it rejects the categorical.

Received November 8, 2000
Revision received October 20, 2003
Accepted October 21, 2003 ■