Reasoners succumb to predictable illusions in evaluating whether sets of assertions are consistent. We report two studies of this computationally intractable task of "satisfiability." The results show that as the number of possibilities compatible with the assertions increases, the difficulty of the task increases, and that reasoners represent what is true according to assertions, not what is false. This procedure avoids overloading memory, but it yields illusions of consistency and of inconsistency. These illusions modify our picture of human rationality.

One view of humans is that they are intrinsically rational. They rely on formal rules of inference similar to those of logic. They sometimes misapply the rules, but haphazardly (1–3). An alternative view is that reasoners construct mental models of what is possible (4–6). Formal rule theories imply that reasoners should infer inconsistency more easily than consistency and should not make systematic errors. The model theory makes the opposite predictions. The results of two studies corroborated the model theory.

The satisfiability problem is intractable because a set of n assertions can be inconsistent even though all its subsets of n − 1 assertions are consistent (7–9), e.g.: If not A then B; if B then C, and not A and not C. Formal rule theories (1–3) have not addressed satisfiability, but they imply that the way to evaluate it is to try to prove the negation of one assertion from the remaining assertions. If successful, the set is inconsistent; otherwise, it is consistent. A single proof establishes inconsistency, but consistency calls for a search for all possible proofs to ensure that none yields the negation of the assertion. Inconsistency should therefore be easier to prove than consistency.

The model theory postulates that reasoners try to construct a mental model for each possibility (6). Thus, an inclusive disjunction: Not-A or B, yields models of the three possibilities (shown here on separate lines):

```
- - a -
- - b -
- - b -
```

They seldom do so spontaneously (4–6) and thus do not notice that the disjunction is equivalent to the conditional, if A then B. Moreover, the conditional has a mental model of the possibility in which the antecedent A is true, but only an implicit model—without explicit content—of the possibilities in which A is false (shown here by an ellipsis):

```
- - a -
- - b -
- - b -
```

One model can show that a set of assertions is consistent, whereas an exhaustive search for models is needed to show that the set is inconsistent. Hence, contrary to formal rule theories, the model theory predicts that consistency should be easier to infer than inconsistency. The task should be easier with conditionals (one explicit model) than with disjunctions (three explicit models). And there should be an interaction. For conditionals, inconsistency should be harder than consistency because the possibilities represented by the implicit model might conflict with another assertion, whereas the difficulty does not arise to the same degree with disjunctions, which have only explicit models. Experiment I tested the three predictions (9). Table 1 presents the percentages of correct responses. The participants were slightly but...
of the disjunction is true, the second clause is false, i.e., there is not both a bolt and a nail, which is incompatible with the second assertion. If the first clause of the disjunction is false, there is neither a pin nor a bolt, which is also incompatible with the second assertion (and the second clause of the disjunction). Hence, the two assertions are inconsistent. When the disjunction occurs with a different second assertion: There is a pin and a bolt, reasoners should evaluate the assertions as consistent, because this assertion matches the third model above. The response is correct, and so these assertions serve as a control problem. Likewise, an illusion of inconsistency and its control are created when other second assertions are used (Table 2).

Experiment 2 tested the occurrence of illusions of consistency and inconsistency, when disjunctive and conditional problems are used (11). Table 2 presents the results of the experiment. The participants tended to succumb to the illusions (12): 128 participants made more errors with illusions than with the controls, and there was one tie—an individual who made no mistakes whatsoever (Sign test, $P = 0.5^{128}$). Similarly, a predicted interaction was highly significant (13): The difference between the illusions of consistency and their controls was larger than the difference between the illusions of inconsistency and their controls (Sign test, $n = 78$, $P < 1$ in 20 million).

Experiment 1 suggested that reasoners assess consistency by envisaging mental models of possibilities. This theory predicts that they should be vulnerable to illusions. Experiment 2 confirmed their occurrence. Indeed, it is striking that individuals go badly wrong with problems based on just three distinct clauses. The illusions are so compelling that they go unnoticed in daily life. If human beings are intrinsically rational, they should make only sporadic errors, similar to slips of the tongue. If they used invalid rules of inference, which would explain the illusions, they would be intrinsically irrational.

But the illusions have no such implication: People understand the explanation of their errors. They lack the capacity to model both truth and falsity. Models of truth alone are a useful compromise. But the compromise does lead to systematic errors in reasoning.

References and Notes
9. The participants were 522 high school graduates in Italy, mean age of about 15 years, who were applicants to the Scuola Superiore Sant’Anna of Pisa, a highly selective Italian university. The experiment was carried out in Italian.
10. The match between the second assertion and the second clause of the disjunction is not essential for the illusion (Table 2), but may enhance it.
11. The participants were a sample of 129 individuals from the same population as in Experiment 1.
12. If the participants interpreted “or else” ("oppure" in Italian) as meaning an inclusive disjunction, then there would be no illusion of consistency in certain cases: "consistent" would be the correct response. A test of native speakers showed that this interpretation is unlikely, and the hypothesis also wrongly predicts that participants will respond "consistent" to illusions of inconsistency and to two of their controls.
13. The model theory predicts that consistent controls should be easier than inconsistent controls, and for the same reason illusions of inconsistency should be less compelling than illusions of consistency. Hence, the difference between illusions of consistency and their controls should be larger than the difference between illusions of inconsistency and their controls.
14. We thank R. Varaldo, director of the Scuola Superiore Sant’Anna of Pisa, for allowing us to test applicants to the school, and S. Pizzini for help in conducting and analyzing the experiments. We also thank three anonymous reviewers and our colleagues R. Byrne, S. Glucksberg, Y. Goldberg, B. Johnson-Laird, H. Neth, M. Newsome, J.-B. van der Heist, and Y. Yang for advice.

23 November 1999; accepted 29 February 2000