How We Detect Logical Inconsistencies

P.N. Johnson-Laird,[1] Paolo Legrenzi,[2] and Vittorio Girotto[2,3]

[1]Department of Psychology, Princeton University; [2]University of Architecture, Venice, Italy; and [3]Laboratoire de Psychologie Cognitive, University of Provence, Aix-en-Provence, France

ABSTRACT—*How do individuals detect inconsistencies? According to the theory described in this article, they search for a possibility represented in a mental model, in which each proposition in a description is true. If they find such a possibility, the description is consistent; otherwise, it is inconsistent. Evidence corroborates the theory. The evaluation of consistency is easy when the first possibility generated from the start of a description fits later propositions in the description; it is harder when this possibility does not fit later propositions, and individuals have to look for an alternative possibility. The theory postulates that models represent what is true, not what is false. As a result, individuals succumb to systematic illusions of consistency and of inconsistency.*

KEYWORDS—*inconsistency; mental models; reasoning*

Do I contradict myself?
Very well then I contradict myself,
(I am large, I contain multitudes.)
–Walt Whitman, *Leaves of Grass*

In Daniel Defoe's novel *Robinson Crusoe*, the hero pulls off his clothes and swims out to his wrecked ship. Soon after he climbs on board, Defoe writes, he fills his pockets with "bisket." But if he has no clothes, how can he have pockets? The two propositions are inconsistent. When enough text separates such details, no one notices their inconsistency. Otherwise, the detection of inconsistency seems trivial. That may be why psychologists have long neglected the topic (but cf. Black, Freeman, & Johnson-Laird, 1986). In this article, we explain why inconsistencies are not trivial, describe a theory of the mental processes underlying their detection, and outline some corroboratory evidence.

THE LOGIC AND PSYCHOLOGY OF INCONSISTENCY

In general, a set of propositions is consistent if there is at least one possibility in which they are all true, and it is inconsistent if they cannot all be true. Inconsistency is devastating in logic, because any proposition whatsoever follows from an inconsistency (e.g., Jeffrey, 1981). Inconsistency is also serious in life, because it is dangerous to believe what is false. It can lead to disaster. As Perrow (1984) remarked about collisions at sea, "Captains still inexplicably turn at the last minute and ram each other. We hypothesized that they built perfectly reasonable mental models of the world, which work almost all the time, but occasionally turn out to be almost an inversion of what really exists" (p. 230). Similarly, the Chernobyl catastrophe was exacerbated by the engineers' failure to believe that the reactor had been destroyed, even though firemen showed them chunks of graphite that they had found. (Graphite is inside a nuclear reactor to moderate its reactions.) The ability to detect inconsistencies is, accordingly, a hallmark of rationality. Psychologists need to understand how people detect them, and what can go wrong in the process.

In the Crusoe example, inconsistency occurs between one proposition and another. But, consider the following set of propositions:

If the reactor is intact, then it is safe.
If the reactor is safe, then no graphite is outside it.
The reactor is intact, and some graphite is outside it.

Together, the three propositions are inconsistent; that is, they cannot all be true. But if any one of the three propositions is dropped, the remaining pair is consistent. A large set of propositions can be inconsistent, but again, if any one of them is dropped, the remaining set is consistent. In general, the detection of inconsistency makes bigger and bigger demands on time and memory as the number of distinct atomic propositions in the set of propositions increases. (An atomic proposition is one that does not contain negation or any sentential connective such as "and" or "or.") These demands can increase so that no feasible computational system, not even a computer that is as big as the universe and runs at the speed of light, could yield a result. A set of, say, 100 atomic propositions allows for 2^{100} possible states of affairs, because each proposition can be either true or false. This number is vast, and, in the worst case, a test of the consistency of the beliefs containing these propositions calls for checking every possibility. If one possibility can be checked in a millionth of a second, it would take more than 40 thousand million million years to examine all the possibilities.

Your beliefs depend on many more than 100 propositions, so how do you maintain consistency among them? One answer is that you keep

Address correspondence to P.N. Johnson-Laird, Department of Psychology, Princeton University, Princeton, NJ 08544.

them segregated into separate sets. Your beliefs about, say, Walt Whitman have nothing to do with your beliefs about golf. In this way, you have some chance of maintaining consistency within sets, though inconsistencies may arise from one set to another. Even with separate small sets of beliefs, the psychological problem remains: How do you determine whether a set is consistent?

A MODEL THEORY OF THE EVALUATION OF CONSISTENCY

One way to evaluate consistency is to rely on formal rules of inference from logic. Here is a typical formal rule:

> If A then B.
>
> A.
>
> Therefore, B.

A and *B* can refer to any propositions whatsoever. Some psychological theories postulate such rules to explain deductive inferences from premises to conclusions (e.g., Rips, 1994). But the evaluation of consistency is not the same task. To evaluate consistency, you need to determine whether a set of propositions can all be true. Nevertheless, formal rules could be adapted to cope with consistency. In this approach, 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, however. We have therefore proposed a different theory based on *mental models*. We refer to this account as the "model" theory.

In brief, the model theory is as follows (see, e.g., Johnson-Laird & Byrne, 1991):

- In reasoning, individuals try to envisage what is possible given the premises and their own knowledge.
- Each possibility that they envisage is represented in a separate mental model.
- A mental model has the same structure as the possibility it represents. For example, a model representing the possibility that Pat owns three cars will have a token representing Pat, three tokens representing cars, and a relation representing ownership holding between the token for Pat and each of the tokens for cars.
- Mental models follow a principle of *truth*: A model represents propositions in the premises only if they are true in the possibility that the model represents. Consider, for instance, the following disjunction about some shapes on a blackboard:

  There is either a circle on the board or a triangle, or both.

  A disjunction is a sentence made up of separate clauses connected by the word "or." The mental models of this disjunction are as follows (each row denotes a model of a separate possibility):

o	
	Δ
o	Δ

  Thus, the first model does not represent that in this possibility, it is false that there is a triangle. The neglect of what is false is usually harmless, but, as we show later, it does yield predictable errors.
- To draw a conclusion from the premises, reasoners find a proposition in their models that is not asserted explicitly in the premises. They check whether it holds in all, some, or a proportion of the models, and they formulate a corresponding conclusion about its necessity, possibility, or probability. To continue the example, suppose individuals are given the following conclusion and asked whether it follows from the premise about the shapes on the blackboard:

  It is possible that there is both a circle and triangle.

  Reasoners can determine that this conclusion does follow from the premise, because the third model corroborates it. They also tend to estimate the probability of both a circle and a triangle being on the board as one third; that is, they assume that each possibility is equiprobable. This account therefore provides an integrated theory of both logical and probabilistic reasoning based on possibilities (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999).
- Reasoners can refute an invalid conclusion using a counterexample, which is a model of the premises in which the conclusion is false.

One assumption extends the theory to deal with consistency (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000):

> Individuals evaluate the consistency of a set of propositions by searching for a model of a possibility in which they are all true. If they find such a model, they evaluate the propositions as consistent; otherwise, they evaluate the propositions as inconsistent.

In an experiment, we asked the participants to consider whether these propositions about what is on a table could all be true at the same time:

1. 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.

Individuals should begin evaluating the consistency of these propositions by constructing a possibility from the first proposition, which is of a form known as a *conditional*. According to the theory (Johnson-Laird & Byrne, 2002), its most salient possibility is

¬ apple	banana

where "¬" denotes negation, "apple" denotes the presence of an apple, and "banana" denotes the presence of a banana. The next step is to use the information in the second conditional to "update" the possibility:

¬ apple	banana	cherry

The third proposition is true in this possibility, and so the set should be judged as consistent.

Now, consider this description:

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

Individuals interpret the first proposition, which is a disjunction, as compatible with three different possibilities, shown here on separate lines:

apple
   banana
apple banana

According to the theory, reasoners should begin by envisaging the first of these possibilities:

apple

The next step is to update this possibility with the first model of the second disjunction:

apple ¬ banana

This possibility, however, is not consistent with the third proposition, and so reasoners should retrace their steps and consider the second model of the first disjunction:

   banana

The second disjunction updates the possibility. The presence of a banana eliminates the possibility corresponding to the disjunction's first clause, and so there must be a cherry:

   banana cherry

When models lack any information about a proposition, individuals tend to interpret this lack as equivalent to the negation of the corresponding proposition. This model lacks an apple, and so individuals infer that there is not an apple, but the model has a cherry, and so it is consistent with the third proposition. However, in contrast to Problem 1, which can be evaluated by generating a single mental model of the propositions, Problem 2 requires finding a second, alternative model. Hence, the evaluation of Problem 2 should be harder than the evaluation of Problem 1.

LIFE IS EASIER WHEN THE FIRST MODEL SUFFICES

Our first experiment tested whether the evaluation of consistency is easier when the first model suffices, as in Problem 1, than when it does not, as in Problem 2 (Johnson-Laird et al., 2000). It exploited the fact that a disjunction implies a conditional. For example, the disjunction

There is an apple or there is a banana.

implies the conditional:

If there is not an apple then there is a banana.

Individuals recognize this implication (Ormerod & Richardson, 2003), though the implication can be blocked when the interpretation of a disjunction or conditional is modified by knowledge (Johnson-Laird & Byrne, 2002). Our experiment contrasted eight sorts of conditional problems (including Problem 1) with eight sorts of disjunctive problems (including Problem 2). We tested 522 of the best high school graduates in Italy, asking them to decide for each problem whether it was possible for the three sentences to be true at the same time. The results showed that the problems based on conditionals had a robust advantage in accuracy (of 15%) over those based on disjunctions, especially when the participants correctly judged that a set was consistent. There was a smaller but significant advantage for consistent problems over inconsistent problems.

Could the results reflect some difference between conditionals and disjunctions other than the nature of their mental models? A subsequent experiment tested the model theory more stringently (Legrenzi, Girotto, & Johnson-Laird, 2003). If the participants judged that a set was consistent, they also had to describe the properties of the corresponding entity. Consider these propositions:

3. 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.

One possibility according to the first two propositions is

saleable elegant stable

The third proposition is true in this possibility, and so individuals should describe the chair as *saleable, elegant, and stable*.

In contrast, consider the following description:

4. 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.

One possibility according to the first two propositions is

unsaleable inelegant unstable

The third proposition is false in this possibility, and so individuals have to search for an alternative model. The first proposition is also compatible with a chair that has these properties:

saleable elegant

The second proposition can be used to update this possibility:

saleable elegant stable

The third proposition is consistent with this model. But the evaluation of consistency should be harder for Problem 4 than for Problem 3 because of the need to reject the initial model. This particular problem might be harder because it contains negatives, and so their occurrence was counterbalanced in the two sorts of problems illustrated in 3 and 4. The results showed that participants were much more accurate when the first model sufficed (97% correct) than when it did not (39% correct).

ILLUSIONS OF CONSISTENCY AND INCONSISTENCY

The principle of truth is central to the model theory. Each simple proposition, affirmative or negative, is represented in a mental model only if it is true in the possibility that the model represents. For instance, consider an *exclusive* disjunction, which allows for one possibility or the other, but not both:

There is an apple or else there is a banana.

This disjunction has two mental models:

apple	
	banana

The first model does not represent explicitly that it is false that there is a banana in this possibility, and the second model does not represent explicitly that it is false that there is an apple in this possibility. Reasoners make "mental footnotes" about what is false, but soon forget them. With these footnotes, however, they can convert mental models into *fully explicit* models:

apple	¬ banana
¬ apple	banana

Hence, as individuals understand, "or else" means that when one proposition is true, the other is false.

A computer implementation of the theory showed that in some cases mental models are wrong about what is possible (Johnson-Laird & Savary, 1999). If reasoners rely on mental models, they should therefore succumb to illusions in cases in which falsity matters. They should evaluate some sets of propositions as consistent when in fact they are inconsistent, and vice versa. Consider this description:

The tray is portable or else not both beautiful and heavy.
The tray is portable and not beautiful.

The mental models of the disjunction in this description are

portable		
	¬ beautiful	heavy
	beautiful	¬ heavy
	¬ beautiful	¬ heavy

The first model includes *portable* but lacks *beautiful*, and so individuals should judge that the second assertion is consistent with it. They would be wrong. People take "or else" to mean that when one proposition is true, the other is false. So if it is true that the tray is portable, then from the disjunctive assertion it is false that the tray is not both beautiful and heavy; that is, it *is* both beautiful and heavy. That is inconsistent with the second assertion. And if it is false that the tray is portable, then that too is inconsistent with the second assertion. Hence, the two assertions are inconsistent.

In contrast, suppose that the same disjunctive assertion occurs with a different second assertion:

The tray is not beautiful and not heavy.

Once again, individuals should judge that the two assertions are consistent (see the fourth of the four possibilities for the disjunction). This time, however, they are correct.

An experiment compared control problems that the theory predicts should yield correct evaluations with experimental problems that the theory predicts should yield illusions, either of consistency or of inconsistency. The results corroborated the theory's predictions: The participants responded more accurately to six sorts of control problems (86% correct) than to six sorts of illusions (27% correct). Only 11 of the 459 participants went against this trend (Legrenzi et al., 2003). Could the participants have misunderstood "or else," failing to realize that when one of its constituent propositions is true, its other constituent proposition is false? A further experiment conveyed its meaning using an unambiguous rubric: "Only one of the following assertions is true." Once again, the participants succumbed to illusions of consistency and to illusions of inconsistency, but responded correctly to the control problems (Legrenzi et al., 2003).

CONCLUSIONS

Social psychologists have known for many years that individuals try to adjust their beliefs to accommodate inconsistencies. In a famous case, members of a cult whose leader had predicted the end of the world reasoned that the prediction had failed as a result of their pious labors (Festinger, Riecken, & Schachter, 1964). But the pioneering work of Kahneman and Tversky (e.g., 2000) showed that individuals are inconsistent in their choices and in their judgments of probabilities. Even experts make the same mistakes when their memory is taxed. Yet, until recently, there were few investigations of how individuals detect inconsistencies. The model theory proposes that they do so when they are unable to accommodate a proposition—an observation or an assertion—into their existing mental models. Two strands of evidence have corroborated this idea. First, in evaluating a set of assertions as consistent, participants are more accurate when the first model of a description suffices than when they must look for a model of an alternative possibility. Second, when falsity matters, they succumb to illusions of consistency and of inconsistency.

Suppose that you are waiting for Paolo, as one of us once was. You believe that he has gone to get the car, and that if he has gone to get the car he will be back shortly. You infer correctly that he will be back shortly. But he does not return even after a quarter of an hour. You detect the inconsistency with the consequence of your beliefs. That is only the first step, though the one that we have tried to explain here. The next step is to modify your beliefs. Psychologists have begun to study this process (e.g., Elio & Pelletier, 1997), but they have no comprehensive account of it. Indeed, you do not just change your beliefs, you try to formulate an explanation that resolves the inconsistency. Perhaps, for example, the car would not start. How you create these explanations is, at present, a mystery.

Recommended Reading

Johnson-Laird, P.N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*, 434–442.

Kahneman, D., & Tversky, A. (Eds.). (2000). (See References)

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (2003). (See References)

Acknowledgments—This research was supported by a grant from the National Science Foundation (Grant 0076287) to the first author to study strategies in reasoning, and by grants from MIUR (Italian Ministry of Universities and Research) and CNR (Italian National Council for Scientific Research) to the other authors.

REFERENCES

Black, A., Freeman, P., & Johnson-Laird, P.N. (1986). Plausibility and the comprehension of text. *British Journal of Psychology*, *77*, 51–62.

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

Festinger, L., Riecken, H.W., & Schachter, S. (1964). *When prophecy fails*. New York: Harper & Row.

Jeffrey, R.C. (1981). *Formal logic, its scope and limits* (2nd ed.). New York: McGraw-Hill.

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

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

Johnson-Laird, P.N., & Savary, F. (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, *71*, 191–229.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge, England: Cambridge University Press.

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

Ormerod, T.C., & Richardson, J. (2003). On the generation and evaluation of inferences from single premises. *Memory & Cognition*, *31*, 467–478.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. New York: Basic Books.

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