Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

Spatial conditionals and illusory inferences

Marco Ragni, Tobias Sonntag & Philip N. Johnson-Laird

To cite this article: Marco Ragni, Tobias Sonntag & Philip N. Johnson-Laird (2016): Spatial conditionals and illusory inferences, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2015.1127925

To link to this article: http://dx.doi.org/10.1080/20445911.2015.1127925

Published online: 28 Jan 2016.

JOURNAL OF COGNITIVE PSYCHOLOGY, 2016
http://dx.doi.org/10.1080/20445911.2015.1127925

Routledge
Taylor & Francis Group

Spatial conditionals and illusory inferences

Marco Ragni[a], Tobias Sonntag[b] and Philip N. Johnson-Laird[c,d]

[a]Department of Artificial Intelligence, Technical Faculty, Freiburg University, Freiburg, Germany; [b]Center for Cognitive Science, Freiburg University, Freiburg, Germany; [c]Department of Psychology, Princeton University, Princeton, NJ, USA; [d]Department of Psychology, New York University, New York, NY, USA

ABSTRACT

Studies of reasoning often concern specialised domains such as conditional inferences or transitive inferences, but descriptions often cut across such domains, for example:

If the circle is to the left of the square then the triangle is to the right of the square.
The square is to the right of the circle.
The triangle is to the right of the square.
Could all three of these assertions be true at the same time?

We report four experiments testing the mental model theory of such problems, which combine spatial transitivity and conditional relations. It predicts that reasoners should try to find a single mental model in which all the assertion hold:

○ □ Δ

Such problems should be easier than those that call for a model in which both clauses of the conditional are false, as when the conditional above occurs with:

The square is to the left of the circle.
The triangle is to the left of the square.

In this case, most participants had the "illusion" that the set was inconsistent (Experiment 1). Analogous results occurred when participants evaluated whether a diagram, such as the one above, depicted a possible spatial arrangement (Experiment 2), and when they evaluated the consistency of a conditional and a conjunction (Experiment 3), and of sets of assertions that contained two conditionals (Experiment 4). The findings appear to be beyond the explanatory scope of theories of reasoning based on logical rules or on probabilities.

ARTICLE HISTORY
Received 22 April 2015
Accepted 22 November 2015

KEYWORDS
Consistency; conditional reasoning; deduction; mental models; relational reasoning; spatial reasoning

Inconsistencies are dangerous. If a description is inconsistent, then it contains at least one false assertion, and to base a decision on a falsehood is a recipe for disaster. As a consequence, individuals try to determine the origins of an inconsistency and to restore consistency (Johnson-Laird, Girotto, & Legrenzi, 2004). The detection of simple inconsistencies is within the competence of naive individuals, that is, those who have not mastered logic, though in general the task is computationally intractable (Cook, 1971). Most psychological theories of reasoning offer no explanation of how individuals assess consistency or of what causes systematic errors in their performance (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000). In contrast, the theory of mental models – the "model theory", for short – answers both these questions, and we describe its answers presently. One potential alternative is that reasoners rely on logical rules of inference (e.g. Rips, 1994). A simple "deductive strategy" is that they evaluate a set as consistent provided that one assertion in the set follows logically from the others. But, this strategy fails in many cases, for example:

The triangle is to the left of the square.
The square is to the left of the circle.

The two assertions can both be true, and so they are consistent with one another. Yet, neither assertion can be deduced from the other. Nevertheless, the evaluation of consistency is closely related to deductive reasoning, and in some formulations of logic, the validity of an inference depends on showing that the negation of its conclusion is inconsistent with its premises (e.g. Jeffrey, 1981). This

CONTACT Marco Ragni ragni@informatik.uni-freiburg.de

relation yields a general deductive procedure for testing consistency: if the *negation* of one assertion in the set follows logically from the remaining assertions, then the set is inconsistent. The procedure seems implausible for naive reasoners, and no previous studies of consistency have ever reported its use (e.g. Johnson-Laird et al., 2004). Indeed, many individuals fail to make the correct negations of assertions (Khemlani, Orenes, & Johnson-Laird, 2013). Rips's system is an advance over its rivals, but, as he proved, it is not a complete inference system (Rips, 1994, Chap. 4), and so it cannot provide a general basis for the evaluation of consistency. Nevertheless, we will examine in the following the simple deductive strategy in which participants try to prove one assertion from the others in order to assess consistency.

Empirical tests of theories of reasoning tend to focus on a particular domain, such as spatial reasoning or conditional reasoning. Problems in daily life, however, often depend on premises that cut across these "text book" domains, for example:

> If the school is to your left, then the bank is to your right.
> You're to the right of the school.
> Therefore, the school is to the left of both you and the bank.

This inference hinges on a conditional relation between spatial arrangements. The model theory applies to conditionals (e.g. Johnson-Laird & Byrne, 2002) and to spatial relations (e.g. Byrne & Johnson-Laird, 1989; Knauff, 2013; Ragni & Knauff, 2013; Ragni, Knauff, & Nebel, 2005). But, how does it fare in explaining inferences that depend on both? The conditional above yields a transitive inference to the conclusion that the school is to the left of the bank, granted that the conditional's *if*-clause is true. Hence, such conditionals are more complex than simple conditionals that do not yield inferences merely from the contents of their two clauses. A well-established phenomenon is that naive individuals have to deliberate in order to make certain deductions from simple conditionals (Johnson-Laird & Byrne, 2002). This difficulty should be exacerbated in the case of conditionals yielding transitive inferences.

No previous psychological experiments appear to have investigated reasoning about conditional relations between spatial arrangements. The aim of the present research was therefore to examine how individuals assess the consistency of sets of assertions containing such conditionals. As the next section shows, the model theory predicts that reasoners should be vulnerable to systematic errors that are sufficiently robust to constitute cognitive illusions. In what follows, the article describes the model theory as it applies to such tasks. It then reports four experiments that corroborate the theory's predictions. Finally, it discusses the implications of these results for other theories of reasoning.

The model theory of spatial descriptions

In outlining the model theory, this section deals first with simple spatial relations, then with conditional relations, and finally with the combination of the two. The model theory of spatial reasoning has been implemented computationally (see Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991, Chap. 9; Knauff, 2013; Ragni et al., 2005; Ragni & Knauff, 2013). The theory and its computer implementations apply to three-dimensional spatial arrangements, but throughout the present paper the layouts are one-dimensional, that is, one shape cannot be on top of another or in front of it. A corollary is that the falsity (or negation) of:

The square is to the right of the circle is equivalent to the affirmative relation:

> The square is to the left of the circle.

A relation such as, _ *is to the right of*_, is transitive in that it yields a valid inference of the following sort:

> *x* is to the right of *y*.
> *y* is to the right of *z*.
> Therefore, *x* is to the right of *z*.

The traditional way to capture transitivity is with a corresponding axiom (or "meaning postulate"):

If *x* is to the right of *y*, and y is to the right of *z*, then *x* is to the right of *z* where *x*, *y*, and *z*, range over the entities in the universe of discourse. This axiomatic approach depends on formal rules of inference (e.g. Rips, 1994). In contrast, the model theory postulates that individuals use the meanings of assertions to make mental simulations (e.g. Khemlani, Mackiewicz, Bucciarelli, & Johnson-Laird, 2013). In the case of spatial assertions, they construct mental models from which transitivity is an emergent property (Johnson-Laird, 1983, p. 205 et seq.; Goodwin & Johnson-Laird, 2005). Consider the following example:

> The circle is the left of the square.
> The square is to the left of the triangle.

They yield the mental model:

○ □ Δ

where the left-to-right axis of the model corresponds to the left-to-right layout of the shapes in the relevant situation. This model supports the conclusion:

The triangle is to the right of the circle

and no other model of the assertions refutes it, and so it is a valid inference.

The controversy between rules and models for spatial reasoning has been resolved experimentally. Sets of premises that call for the same formal proof, but that differ in the number of spatial layouts to which they refer, differ in difficulty. Some descriptions are spatially indeterminate, for example:

The square is to the right of the circle.
The circle is to the left of the triangle.

They are consistent with two distinct layouts, both of which are possible:

○ Δ □

and:

○ □ Δ

Experiments have shown that reasoning from indeterminate descriptions is more difficult than reasoning from determinate descriptions, even when an identical formal proof in both conditions yields the required conclusion (e.g. Byrne & Johnson-Laird, 1989). The same result holds for reasoning about temporal relations (e.g. Schaeken, Girotto, & Johnson-Laird, 1998; Schaeken, Johnson-Laird, & d'Ydewalle, 1996a, 1996b). Likewise, experiments have corroborated the model theory of ternary relations (Jahn, Knauff, & Johnson-Laird, 2007), interval relations (Knauff, 2013), topological relations (Knauff & Ragni, 2011), and relations between relations (Goodwin & Johnson-Laird, 2006). Hence, models are better than rules in explaining spatial reasoning.

Conditionals are much more controversial. According to the model theory, the meanings of conditionals and of other connectives, such as: *or*, *and*, and *before*, refer to sets of possibilities (Johnson-Laird & Byrne, 2002). The conditionals in the present studies are *basic* in that the meaning of *if_then_* is not affected by the contents of the *if*-clause and the *then*-clause, by context, or by general knowledge. Basic conditionals have two mental models, which are shown here on separate lines:

A B
…

The first mental model is the only explicit one and it represents the possibility in which *A* and *B* both hold. The second model denoted by the ellipsis has no explicit content, and it represents the possibilities in which *A* does not hold. On this account, a pair of conditionals, *If A then B,* and *If A then not B,* appear to be inconsistent because of the conflict between their explicit mental models.

The mental models of a basic conditional, *If A then B*, can be fleshed out into a set of *fully explicit* models corresponding to three different possibilities (Johnson-Laird & Byrne, 2002):

A	B
not-A	not-B
not-A	B

where negation (not) is used to represent false propositions. The theory postulates that the models are fleshed out in the order shown here. And, as Barrouillet and his colleagues have shown, it corresponds to the order in which children acquire these possibilities (see e.g. Barrouillet, Grosset, & Lecas, 2000; Barrouillet & Lecas, 1999). In terms of a conditional in which *A* is "the circle is to the left of the square" and *B* is "the square is to the left of the triangle", the meanings of these possibilities correspond to the following one-dimensional layouts of the shapes, which reflect an indeterminacy in the *not-A* and *B* contingency:

○	□	Δ	(A	B)
Δ	□	○	(not-A	not-B)
□	○	Δ	(not-A	B)
□	Δ	○	(not-A	B)

Of course, most people are highly unlikely to think of these four possibilities, at least spontaneously. They are more likely to rely on mental models.

The distinction between mental models and fully explicit models is a special case of a "dual process" theory (see e.g. Evans, 2007; Evans & Stanovich, 2013; Kahneman, 2011; Stanovich, 1999). Earlier intimations of dual processes in reasoning are due to Wason and Johnson-Laird (1970). The model theory embodied them from its origins, postulating that intuitions rely on a single mental model, which is constructed without access to working memory (Johnson-Laird, 1983, Chap. 6), whereas fully explicit models call for access to working

memory and therefore processes that have more computational power. When tasks are easy, such as listing the possibilities to which conditionals refer, individuals can use this greater power to construct fully explicit models. This distinction between System 1 for intuitions and System 2 for deliberations that yield alternative models has been modelled in a computer programme, *mReasoner* (Khemlani & Johnson-Laird, 2013), which has been extended to deal with the distinction between intuitions and deliberations in inferring probabilities (Khemlani, Lotstein, & Johnson-Laird, 2015). The crucial difference in computational power is that System 1 can construct only a single representation at a time and can run loops of operations for only a small number of times, whereas System 2 can construct alternative models by fleshing mental models out and carry out loops of operations until it exhausts the processing capacity of an individual's working memory. No other dual process theory, as far as we know, has been modelled in a computer programme.

The model theory postulates that content and general knowledge can modulate the process of interpretation. They can prevent the construction of models of certain possibilities. They can yield, for example, a biconditional interpretation equivalent to: *If and only if A, then B,* as in *If it rained then it poured* (Johnson-Laird & Byrne, 2002, p. 663), which is compatible with only two fully explicit possibilities, because it cannot pour without raining:

Rained	poured	(A B)
Not-rained	Not-poured	(Not-A Not-B)

They can also establish temporal and other relations between *A* and *B* (e.g. Johnson-Laird & Byrne, 2002; Juhos, Quelhas, & Johnson-Laird, 2012). It follows that the mental algorithm for interpreting conditionals cannot be truth-functional. The process of interpretation yields models of possibilities, not the truth values of clauses. Conversely, if the interpretative system had access only to the truth values of the *if-clause* and the *then*-clause, then it would be unable to take into account temporal relations or to determine which sets of possibilities a conditional refers to. The possibilities represented in these fully explicit models of a basic conditional, *If A then B*, correspond to the true rows of a truth table for material implication in sentential logic. But, possibilities are not truth values. A truth table represents a disjunction of alternative cases, whereas a basic conditional refers to a set of possibilities, and sets have the force of conjunctions (Johnson-Laird, Khemlani, & Goodwin, 2015).

Individuals can judge that a set of assertions is consistent if they can envisage a possibility common to them all, because then they could each be true (see Jeffrey, 1981, p. 8). The model theory accordingly postulates that individuals evaluate consistency by trying to construct a model of such a possibility. If they can find such a model, then they judge that the set is consistent; otherwise, they judge that is inconsistent. As we now illustrate, the way in which they carry out this process is to interpret the assertions in the set, clause by clause.

A typical problem from our experiments (problem 2, Experiment 1) is as follows:

> If the square is to the right of the circle then the triangle is to the right of the square.
> The circle is to the left of the triangle.
> The square is to the left of the triangle.
> Could all three of these assertions be true at the same time?

The problem, in effect, calls for an evaluation of the consistency of the three assertions, and the correct evaluation depends both on the conditional and on the two categorical assertions. The first clause of the conditional above yields the model:

○ □

and the second clause updates this model to:

○ □ Δ

As before, the left-to-right axis in this diagram corresponds to the left-to-right layout of the shapes in the relevant situation. The second assertion in the example, *the circle is to the left of the triangle*, holds in this model – indeed, the model yields this transitive conclusion. And the third assertion, *the square is to the left of the triangle*, also holds in the model. The three assertions are therefore consistent with each other, and so they could all be true at the same time. In contrast, the deductive strategy leads individuals to infer that a set is consistent if they can deduce one of its assertions from the others. Hence, the strategy cannot explain this evaluation, because no two assertions in the set logically imply the third assertion. There is no need to go beyond the single mental model of the conditional in order to determine that the assertions are consistent.

A contrasting problem from our experiments (problem 4, Experiment 1) is as follows:

If the circle is to the left of the square then the square is to the left of the triangle.

The circle is to the right of the triangle.
The circle is to the right of the square.

The process of constructing a mental model of the conditional proceeds as before, and yields:

○ □ Δ

But, the second assertion does not hold in this model, and neither does the third assertion. Hence, reasoners should tend to respond that the three assertions cannot all be true at the same time. This evaluation, however, is an illusion. Reasoners can avoid it only if they flesh out their mental models of the conditional into a fully explicit model of the conditional in which its *if*-clause does not hold and its *then*-clause does not hold:

Δ □ ○

Both the second assertion and the third assertion in the set hold in this model, and so the three assertions can in fact all be true at the same time contrary to the illusory evaluation.

As these contrasting examples show, the model theory's predictions hinge on the distinction between System 1, which uses a single mental model, and System 2, which can use fully explicit models. Hence, the first sort of problem above, which can be solved from a single mental model of the conditional, should be easier than the second sort of problem, which can be solved only by constructing at least one additional fully explicit model. The theory therefore predicts that evaluations of consistency should be likely to be correct when mental models yield them but to yield illusions of inconsistency when they depend on fully explicit models. We show later that evaluations of consistency can differ in predicted difficulty too.

Experiment 1

The participants' task in this experiment was to evaluate the consistency of sets of assertions. Because naive individuals are sometimes confused by the concept of "consistency", we framed the problems in the equivalent terms of whether or not it was possible for the assertions in a set to all be true at the same time. Each set contained one conditional assertion and two categorical assertions, and so a typical problem (problem 1) was:

If the circle is to the left of the square, then the square is to the left of the triangle.

The square is to the right of the circle.
The triangle is to the right of the square.

Could all three of these assertions be true at the same time?

For convenience, we abbreviate problems throughout the paper using the following notation:

If A then B
′A
′B

where *A* relates *x* to *y, B* relates *y* to *z,* and *C* relates *x* to *z*. ′*A* stands for the converse of the relation *A*, and so if *A* is true then so is ′*A*. The problem can be solved from a mental model alone. A contrasting problem (problem 4) has the following structure:

If A then B.
¬B.
¬A.

where ¬*B* and ¬*A* denote implicit negations of relations, for example, if *A* denotes *x is to the left of y,* then ¬*A* denotes *x is to the right of y* or its equivalent converse: *y is to the left of x*. Granted the one-dimensional nature of the problems, *A* is true if, and only if, ¬*A* is false. An example of problem 4 is accordingly:

If the circle is to the left of the square
then the square is to the left of the triangle.
(If A then B)
The triangle is to the left of the square. (¬B)
The square is to the left of the circle. (¬A)

The mental models of the problem suggest that the assertions cannot all be true at the same time, but the fully explicit models of the conditional show on the contrary that all three assertions can be true at the same time (see the previous section of the paper).

The experiment examined two consistent problems that could be solved using mental models (problems 1 and 2), and two consistent problems that could be solved only using fully explicit models (problems 3 and 4). Both sorts of problem had matching inconsistent problems in which the third assertion did not hold in either the mental models or the fully explicit models of the previous assertions. All possible orders of the shapes occurred in the experiment, and participants were therefore forced to interpret the conditionals and the categoricals in order to arrive at a principled evaluation of the assertions. The model theory predicts that problems that can be solved using mental models

should be easier than those that can be solved only using fully explicit models, and that the latter problems should yield illusory evaluations that the sets of assertions are inconsistent.

Participants

The experiment tested 24 participants (12m/12f, mean age 28 years) on Amazon's Mechanical Turk. They received a nominal fee for their participation. We took the usual precautions for this procedure, for example, the programme checked that participants were native speakers of English, and it allowed only one participant from a given computer.

Materials

Table 1 below presents the eight spatial descriptions used in the experiment, which each consisted of a conditional and two categorical assertions. The third assertions were either consistent or inconsistent with the two previous assertions in the descriptions, and so half the problems were consistent and half were inconsistent. In the experiment, the spatial relations assigned to *A*, *B*, and *C* were *x is to the left of y*, *y is to the left of z*, and *x is to the left of z*; where the lowercase letters *x*, *y*, and *z*, denote different shapes. We emphasise that *x*, *y*, and *z*, are variables and so the particular shapes assigned to them varied from one problem to the next for each participant over the course of the experiment. The six possible assignments of the shape terms to problems occurred in the experiment for each of the participants.

Design

Each participant carried out 2 practice trials, and then 4 problems in each of the 8 conditions illustrated in Table 1 below, that is, a total of 32 problems in the experiment proper. The four instances of a given problem had different assignments of the three shape terms. The order of the problems was randomised for each participant.

Procedure

The participants were told that the task was not a test of intelligence or personality. They would read sets of three assertions, and for each set they had to answer the question: "Could all three of these assertions be true at the same time?" They were told that they were not allowed to make notes or to draw diagrams, but should try to solve each problem mentally. We used the separate-stage paradigm (Potts & Scholz, 1975): the conditional assertion and the second assertion were presented simultaneously. The participants then pressed the space bar to see the third assertion and the previous two assertions disappeared. They responded by pressing one of two buttons on the screen, one labelled "yes" and one labelled "no". They could take as much time as they needed, but they had to try to answer correctly. The programme recorded their evaluations and latencies to respond from the onset of the first pair of assertions. In this experiment and the subsequent ones, we used a flash implementation in MTurk, and so the participants' evaluation times were recorded on their computers and then sent to us. Hence, the evaluation times did not depend on any transmission time. Before we carried out any of the experiments, we made extensive tests of this MTurk system to ensure that the evaluation times were reliable.

Results and discussion

Table 1 presents the results for the eight sorts of problem. The participants solved 65% of the problems correctly. As predicted, the consistent problems were reliably easier when a mental model satisfied all the assertions (74% correct for problems

Table 1. The four consistent problems and the four inconsistent problems in Experiment 1: the mental models of the first two assertions satisfy the third assertions in the consistent versions of problems 1 and 2 but not in the consistent versions of problems 3 and 4, and so, as the arrow shows, their correct evaluation depends on finding a fully explicit model of the conditional in which each of its clauses is false.

		Third assertion	
The first two assertions	Model of the first two assertions	Consistent	Inconsistent
1. If A then B. ′A	A B	′B. 79	¬ B. 73
2. If ′A then′B. C	′A′ B	B. 69	¬ ′B. 70
3. If ′A then B. ¬ B	′A B ⇒ ¬ ′A ¬ B	¬ ′A. 43	A. 77
4. If A then B. ¬ C	A B ⇒ ¬ A ¬ B	¬ A. 35	′A. 78

Notes: The right-hand columns present the percentages of correct evaluations, that is, "yes" responses to consistent assertions and "no" responses to inconsistent assertions. The abbreviation *A* stands for *x is to the left of y*, *B* stands for *y is to the left of z*, and *C* stands for *x is to the left of z*. ′*A* stands for the converse of a relation, which is true too, and ¬*A* stands for its implicit negation, so that when one is true the other is false.

1 and 2) than when a fully explicit model was necessary to satisfy all the assertions (39% correct for problems 3 and 4; Wilcoxon test, $z = 3.18$, $p < .001$, $r = .46$). We used non-parametric statistical tests in all of our analyses in order to obviate problems of distribution, and also to allow us to examine the reliability of predicted ordinal trends (using Page's L test in Experiments 3 and 4). These tests assess a stochastic increase from one condition to another, so they are less powerful than parametric tests, such as analysis of variance, and less likely to lead to an incorrect rejection of the null hypothesis (a Type I error). The participants were not evaluating the consistent problems at random, because 19 out of the 24 participants performed better than chance on problems 1 and 2, and there were 4 ties (Binomial test, $p < .000025$), whereas 15 participants did worse than chance with problems 3 and 4, and there was one tie (Binomial test, $p < .0025$). This pattern of results corroborates the use of mental models in the evaluations, because they should lead to illusory evaluations of inconsistency for problems 3 and 4, which can be evaluated correctly only from fully explicit models.

The inconsistent problems did not differ reliably in difficulty, which was to be expected because in all of them the third assertion was inconsistent with any mental model (or fully explicit model) satisfying the previous assertions. They were easier than the consistent problems (74% vs. 56%, Wilcoxon test, $z = 2.89$, $p < .01$, $r = .42$), because the difficulty of the consistent problems depended in part on fully explicit models for their solutions. The latencies of evaluations did not differ reliably among the different sorts of problems, and so we have spared readers a detailed analysis. But, the consistent problems were not correctly evaluated reliably faster when mental models sufficed (21.5s, SE 1.6) than when fully explicit models were called for (20.0s, SE 4.4; Wilcoxon test, $z = .88$, $p > .35$). In two of the consistent problems, the second assertion referred to the same items as one of the clauses in the conditional (problems 1 and 3), and in two of the consistent problems, the second assertion referred to one item in the *if*-clause and one item in the *then*-clause (problems 2 and 4). In this second case, reasoners need to construct a complete model of the transitive relation in order to accommodate the second assertion whereas it is not strictly necessary to do so for the first case. The factor had an additional effect on the difficulty of the problems (61% vs. 52%; Wilcoxon test, $z = 1.69$, $p < .05$, $r = .24$).

One putative alternative explanation of the results, which we described earlier, is the deductive strategy in which individuals evaluate a set of assertions as consistent if they can deduce one member of the set from the others. Problems 2 and 4, however, cannot be evaluated in this way, because no assertion in them can be deduced from the others. Problem 2 was easier than problem 4 (see Table 1), and the deductive strategy cannot account for the difference. Another possibility is that participants use some sort of "matching" strategy. Consider, for instance, an instance of problem 1:

(a) If the circle is to the left of the square, then the triangle is to the right of the square.
(b) The square is to the right of the circle.
(c) The triangle is to the right of the square.

Participants could match (c) to the *then*-clause of (a) and detect that (b) is the converse of the *if*-clause of (a). Such a strategy, however, fails for problem 2, in which participants need to make a transitive inference. This problem is 10% harder than problem 1, but the matching strategy fails to explain the overall level of performance with problem 2 (69% correct).

The model theory explains all the results. The participants performed better than chance when mental models sufficed for an evaluation of consistency but worse than chance when fully explicit models were required to do so. The task was harder when models had to embody a transitive relation than when they did not have to.

Experiment 2

To what extent do the results of Experiment 1 depend on the use of a third assertion, which was either consistent or inconsistent with the previous two assertions in a description? According to the model theory, a diagram in place of an assertion should yield similar but superior evaluations. Diagrams in general do not guarantee enhanced reasoning, but when they make possibilities salient, they map directly into mental models and eliminate the need for linguistic processing. The result is faster and more accurate inferences (Bauer & Johnson-Laird, 1993). Experiment 2 tested this prediction for problems, such as:

If the circle is to the left of the square, then the square is to the left of the triangle.

The square is to the right of the circle.
Is this a possible layout?

○ □ Δ

Theories based on formal rules of inference, including the deductive strategy, offer no account of such inferences, because the problems introduce the modal notion of possibility, and depend on diagrams, which both lie outside their scope. As before, the experiment tested the difference between problems solvable with mental models and problems solvable only with fully explicit models. Table 2 below presents the 12 sorts of problem using the same conditional in order to help readers to grasp the design, though the conditional varied from one trial to another in the actual experiment.

Participants

We tested 31 new participants (17m/14f, mean age 31 years) from the same population as before.

Design, materials, and procedure

The problems were analogous to those of Experiment 1 but with diagrams instead of third assertions, and the evaluation of possible layouts rather than of the consistency of assertions. There were 2 practice trials and 12 sorts of problem: 6 consistent problems and 6 inconsistent problems. Three sorts of consistent problem had a mental model of the two assertions that satisfied the diagram (problems 1, 2, and 3, in Table 2), and three sorts of consistent problem had only a fully explicit model that satisfied the diagram (problems 4, 5, and 6 in Table 2). The categorical assertion was the converse of the *if*-clause (problem 1 in Table 2 below), the converse of the *then*-clause (problem 2), or the converse of a transitive relation between entities referred to in the two separate clauses (problem 3). The same manipulation occurred for the fully explicit model problems (see problems 4, 5, and 6, respectively). The method also enabled us to use problems 2 and 4, which could not have been used in Experiment 1. The experiment tested 4 different instances of consistent mental model problems (12 problems in all), 4 different instances of the consistent fully explicit model problems (12 problems in all), and 4 different instances of the inconsistent problems (24 problems in all). There were accordingly 48 different problems in the experiment as a whole. The 6 possible assignments of the 3 shapes to the variables in Table 2 occurred roughly equally often in the experiment for each participant.

Each of the 6 sorts of consistent problem, and each of the 6 sorts of inconsistent problem, was presented 4 times with different assignments of the 3 shapes, for a total of 48 trials. Each participant received the 48 problems in a different randomised order. The procedure was identical to that of Experiment 1 except that the task was to evaluate the possibility of a spatial arrangement depicted in a diagram instead of the possible truth of assertions.

Results and discussion

Table 2 presents the percentages of correct evaluations for each of the problems in the experiment. The participants made 76% correct evaluations. As predicted, the consistent problems were easier when a mental model of the assertions satisfied the diagram (88% correct for problems 1, 2, and 3) than when only a fully explicit model of the assertions satisfied the diagram (42% correct for problems 4, 5, and 6; Wilcoxon test, $z = 3.98$, $p < .001$, $r = .51$). The inconsistent problems did not differ reliably in difficulty, which was expected because

Table 2. The six consistent problems and the six inconsistent problems in Experiment 2: the mental models of the pairs of assertions satisfy the layout in the consistent versions of problems 1, 2, and 3, whereas, as "⇒" signifies, they need to be fleshed out into a fully explicit model in the consistent versions of problems 4, 5, and 6.

		Presented layout	
The first two assertions	Models of the first two assertions	Consistent	Inconsistent
1. If A then B. ′A	A & B	A & B 91	¬A & ¬B 87
2. If A then B. ′B	A & B	A & B 88	¬A & ¬B 88
3. If ′A then′B. C	A & B	A & B 85	¬A & ¬B 86
4. If A then′B. ¬A	A & B ⇒ ¬A & ¬B	¬A & ¬B 42	A & B 87
5. If ′A then B. ¬B	A & B ⇒ ¬A & ¬B	¬A & ¬B 40	A & B 88
6. If A then B. ¬C	A & B ⇒ ¬A & ¬B	¬A & ¬B 42	A & B 89

Notes: The right-hand columns present the two layouts that the participants had to evaluate: one consistent and the other inconsistent with the assertions. These columns also present the percentages of correct evaluations in the experiment. *A* denotes a relation between *x* and *y*, B denotes a relation *y* and *z*, and C denotes the transitive relation between *x* and *z*, where the lower-case letters *x*, *y*, and *z*, stand for different variables to which the three shapes were assigned at random. If *A* denotes the relation *x is to the left of y*, then ′*A* denotes its equivalent converse, *y is to the right x*, and ¬*A* denotes its implicit negation, *x is to the right of y*.

in all of them the diagram was inconsistent with any models of the assertions. They were easier than the consistent problems (87% vs. 65%, Wilcoxon test, $z = 3.73$, $p < .001$, $r = .47$), in part because of the difficulty of the problems depending on fully explicit models.

As in the previous experiment, the participants were not evaluating the consistent problems at random, because 28 out of the 31 participants performed better than chance with those problems that could be solved using mental models, and there were no ties (Binomial test, $p < .001$), whereas 18 participants did worse than chance with problems that could be solved using only fully explicit models, there were no ties, and so this difference was not reliable (Binomial test, n.s., $p = .437$). We rejected the latency data from one participant who had one trial with a latency of 84 s. Unlike the previous experiment, the consistent problems were correctly evaluated reliably faster when mental models sufficed (9.8s, SE 0.45s) than when fully explicit models were called for (11.7s, SE 2.5s; Wilcoxon test, $z = 2.04$, $p < .025$; $r = .42$).

The use of a diagram of a layout should in principle improve performance in comparison with three assertions in the previous experiment. Performance was indeed better in the present experiment (76% correct) than in the previous experiment (65% correct) in a by materials analysis of the problems in common to the two experiments (Wilcoxon test, $z = 4.28$, $p < .001$; $r = .58$). It is not proper to compare the results of experiments, which may differ in too many respects for a sensible comparison, but we report the result because it challenges theories that do not rely on mental models. Existing theories based on logical rules have not been framed to account for how people cope with diagrams, but, because these theories use rules that manipulate linguistic expressions, they predict that the present task should be more difficult than one in which conclusions are assertions. In fact, as the model theory predicts, prior results show that reasoning with diagrams that make possibilities explicit is easier than reasoning with equivalent assertions (Bauer & Johnson-Laird, 1993).

Experiment 3

The previous experiments examined sets of assertions containing a single conditional, and the difficult problems to evaluate were those in which a categorical assertion or diagram referred to a possibility in which both clauses of the conditional were false. Hence, they could be solved correctly only from a fully explicit model. In contrast, the present experiment examined pairs of assertions in which one assertion was a conditional and the other assertion was a conjunction. Given a conditional, *If A then B*, it should be easy to determine its consistency with a conjunction, *A and B*, and easy to determine its inconsistency with a conjunction, *A and* ¬*B*, where ¬*B* denotes an implicit negation, for example, if *B* denotes *y is to the left of z*, then ¬*B* denotes either *y is to the right of z* or *z is to the left of y*. Both the conjunctions, ¬*A and* ¬*B* and ¬*A and B*, are consistent with the conditional, but reasoners need to consider fully explicit models of the conditional in order to respond correctly to these assertions. Here is an example of a problem of the form: *If A then B*; ¬*A and B*:

> If the circle is to the left of the square, then the square is to the left of the triangle.
> The circle is to the right of the square and the square is to the left of the triangle.
> Could both these assertions be true at the same time?

According to the model theory, participants should form a mental model of the conditional: *A B*, and test if it matches the mental model of the conjunction. There is no match in this case, because the second assertion has the model: ¬*A B*. Individuals should tend to rely mainly on mental models, but to the extent that they do flesh out models explicitly, they should show a reliable trend in increasing difficulty over the three sorts of conjunction that are consistent with the conditional: *A and B*, ¬*A and* ¬*B*, ¬*A and B*. The third of these conditions is a case in which the deductive strategy fails to establish consistency. Given a conditional, *If A then B*, ¬ *A* does not yield *B* as a valid conclusion, and *B* does not yield ¬*A* as a valid conclusion. Yet, the two assertions are consistent.

Participants

The experiment tested 40 participants (20m/20f; mean age: 33 years) from the same population as before.

Design, materials, and procedure

The problems were analogous to those in Experiment 1. They consisted of a conditional assertion and four sorts of conjunction of spatial relations; *A and B*, *A and* ¬*B*, ¬*A and B*, and ¬*A and* ¬*B*. In half

of the problems the conjunction was presented first and then the conditional, and in the other half of the problems, the conditional was presented first and then the conjunction. There were a total of 24 problems: 12 consistent problems (4 of each of the 3 sorts) and 12 inconsistent problems. All inconsistent problems were of the form: *If A then B*, *A and* ¬*B*, for example:

If *x* is to the left of *y* then *y* is to the left of *z*.
x is to the left of *y* and *z* is to the left of *y*.

The participants carried out all the problems shown in Table 3 below. The experiment systematically varied the spatial relations, and included two versions of each of the four sorts of problem:

If A then B.
A and B.

and:

If ′A then ′B.
′A and ′B.

In this way, the design ensured that the 12 inconsistent problems were all different. Each of the problems was presented with different assignments of the 3 shapes, for a total of 24 trials. The six possible assignments of the three shapes to the variables in Table 3 below occurred roughly equally often in the experiment for each participant, who each received the problems in a random order. The general procedure was the same as that for Experiment 1.

Results

The participants made 60% correct evaluations. The order of the two assertions in the pair had no reliable effect on either accuracy or latency: 61% correct, mean latency of 20.5s, SE 1.2s when the conditional was first; and 59% correct, mean latency of 23.8 s, SE 1.5 s when the conjunction was first (Wilcoxon test, $z = 0.67$, $p = .51$, ns, and $z = 1.38$, $p = .17$, ns, respectively). Hence, we pooled the results for the subsequent analyses. Table 3 presents the percentages of correct evaluations for the four sorts of problem, and the mean latencies for all the evaluations, whether correct or incorrect. We used all the latencies because of the small percentages of correct evaluations in two conditions. The difficulty of the consistent problems showed reliable stochastic rank-order trends both in accuracy – despite the same 15% accuracy for the two more difficult conditions (Page's L test, $z = 4.42$, $p < .00001$) – and in the latency of all evaluations (Page's L test, $z = 3.47$, $p < .001$). As in previous experiments, the consistent problems that could be evaluated correctly from mental models yielded a greater than chance accuracy, whereas those that could be evaluated correctly only from fully explicit models yielded worse than chance accuracy (Wilcoxon test, $z = 4.86$, $p < .0001$; $r = .54$). Only 5 out of the 40 participants performed at chance or better with problems calling for fully explicit models. This result suggests that with complex conditionals only a minority of individuals can engage System 2.

Experiment 4

The previous studies examined sets of assertions containing single conditionals. The aim of our final experiment was to test whether the model theory's predictions applied to sets containing two conditionals, including some conditionals that are spatially indeterminate and accordingly have two mental models with explicit contents. The theory distinguishes three sorts of consistent problem. The first and theoretically easiest problem (I_1 in Table 4) is one in which a single mental model satisfies all three assertions, for example:

If the circle is to the left of the square then the square is to the left of the triangle.
If the square is to the left of the triangle then the circle is to the left of the square.

Table 3. The three sorts of consistent problem and the one sort of inconsistent problem in Experiment 3, the percentages of correct evaluations, and the mean latencies in s for all evaluations, correct or incorrect (SE's in parentheses).

			The four categorical assertions, their consistency or inconsistency with the conditional, and the percentages of correct evaluations, overall mean latencies in s and standard errors in parentheses			
The conditional	Mental models of the conditional	Fully explicit models of the conditional	1. A & B	2. ¬A & ¬B	3. ¬A & B	4. A & ¬B
If A then B	A & B …	A & B ¬A & ¬B ¬A & B	Consistent 79 17.93 (1.6)	Consistent 15 20.67 (1.57)	Consistent 15 23.3 (2.0)	Inconsistent 84 22.0 (1.2)

Notes: The abbreviation *A* denotes *x is to the left of y, B* denotes *y is to the left of z,* and ¬ stands for an implicit negation, for example, ¬ A denotes *x to the right of y*, where *x, y, z,* denote variables to which the three shapes were assigned at random.

Table 4. The 16 problems in Experiment 4 based on four sorts of pairs of conditionals and four sorts of third assertion with the mental models of each conditional. Each cell states whether the set of assertions is consistent or inconsistent and the number of conditionals with mental models in which the third assertion holds: 2, 1, or 0.

			The third assertion, its consistency or inconsistency with the conditionals, the number of conditionals with mental models in which it holds, and the percentages of correct evaluations			
	The two conditionals	Mental models of each conditional	1. A & B	2. ¬A & ¬B	3. A & ¬B	4. ¬A & B
I.	If A then B	A & B …	Consistent: 2 98	Consistent: 0 22	Inconsistent: 0 87	Inconsistent: 0 82
	If B then A	A & B …				
II.	If A then B	A & B …	Consistent: 1 47	Inconsistent:0 88	Inconsistent: 1 77	Consistent: 0 15
	If ¬B then A	A & ¬B …				
III.	If A then B	A & B …	Inconsistent: 1 53	Consistent: 0 20	Inconsistent:0 88	Consistent: 1 42
	If B then ¬A	¬A & B …				
IV.	If A then B	A & B …	Consistent: 1 45	Consistent: 1 43	Inconsistent:0 90	Consistent: 0 15
	If ¬B then ¬A	¬A & ¬B …				

Notes: If *A* denotes *x is to the left of y,* then *B* denotes *y is to the left of z, and* ¬*A* stands for an implicit negation of *A*, that is, *x to the right of y,* and, where *x*, *y*, *z*, denote variables to which the three shapes were assigned at random. The second conditional for half the problems of each of the four sorts used the equivalent converse relations to those in the first conditional.

The circle is to the left of the square, which is to the left of the triangle.
Could all three of these assertions be true at the same time?

Its structure is as follows:

If A then B.
If B then A.
A & B.

where *A* stands for *x is to the left of y*, and *B* stands for *y is to the left of z*. Here *A & B* designates an assertion, as in the example above, which is equivalent to a conjunction. In the example, a mental model of the first conditional satisfies both of the subsequent assertions, and so its evaluation should be easy. We refer to this sort of problem as "consistent 2", because a mental model of the two conditionals satisfies all the assertions.

The second sort of problem (e.g. IV_1 in Table 4) is:

If A then B.
If ¬B then ¬A.
A & B.

In this case, the first conditional yields a mental model such as:

○ □ Δ

whereas the second conditional yields a conflicting mental model:

Δ □ ○

But, the second conditional has a fully explicit model in which both its clauses are false, and this model is the same as the mental model of the first conditional, and it also satisfies the third assertion. The problem has only one conditional with a mental model satisfying the third assertion, and we refer to such problems as "Consistent 1" problems. They should be more difficult than the previous sort of problem, because participants need to flesh one conditional with a fully explicit model in order to respond correctly.

The third sort of problem (e.g. I_2 in Table 4) is:

If A then B.
If B then A.
¬A and ¬B.

The mental models of neither conditional satisfy the third assertion. But, if each conditional is fleshed out with a fully explicit model, they each have a model in which both of their clauses are false, and this model satisfies the third assertion. Because neither conditional has a mental model satisfying all the assertions, we refer to such problems as "consistent 0" problems. They should be the most difficult of all, because both conditionals need to be fleshed out with a fully explicit model in order to yield a principled evaluation. In sum, the theory predicts the following trend of increasing difficulty for consistent problems: the third assertion holds in a mental

model of the two conditionals (consistent 2), it holds in a mental model of one of the conditionals (consistent 1), and it holds in a mental model of neither of the conditionals (consistent 0).

Inconsistent problems should also differ in difficulty in this experiment. By definition, the third assertion holds in neither the mental models nor the fully explicit models of the two conditionals. But, in some cases, a mental model of one of the conditionals does satisfy the third assertion (inconsistent 1), whereas in other cases, no mental model of a conditional satisfies the third assertion (inconsistent 0). The former problem should be harder than the latter to evaluate correctly as inconsistent. In the former case (inconsistent 1), the match with one conditional may elicit an erroneous judgment of consistency. The present experiment tested these predictions about consistent and inconsistent sets of assertions.

Participants

We tested 32 participants (14m/18f, mean age 36 years) from the same population as before.

Design, materials, and procedure

The sets of assertions were based on four sorts of pairs of conditionals according to the following scheme:

I	II	III	IV
If A then B. If B then A.	If A then B. If ¬B then A.	If A then B. If B then ¬A	If A then B. If ¬B then ¬A.

where, as usual, ¬*A* and ¬B denote implicit negations. For all four sorts of problem, the second conditionals were of two sorts, either with the same relations as the first conditional or else with their equivalent converses. Each pair of conditionals was combined in separate sets of assertions with four sorts of third assertion, which were equivalent to the following: *A & B*, ¬*A* & ¬*B*, *A* & ¬*B*, and ¬*A* & *B*. These conjunctions were expressed using the following form of words:

x is to the left of *y*, which is to the left of *z*.

The values of *x*, *y*, and *z*, in the experiment were again the geometrical shapes: triangle, square, and circle. Table 4 below presents the resulting 16 sorts of problem. The table shows instances of the 3 sorts of consistent problem (consistent 2, 1, and 0) in which the third assertion matches the mental models of both the two conditionals (problem I_1), of one conditional (problems II_1, III_4, IV_1, IV_2), and of neither conditional (problems I_2, II_4, III_2, IV_4). Likewise, it shows the two sorts of inconsistent problem (inconsistent 0 and 1) in which the third assertion matches none of the models of the conditionals (problems I_3, I_4, II_2, III_3, IV_3), or matches the mental models of one of the conditionals II_3, III_1). There were four practice problems in a fixed order based on a single conditional. The 16 problems, 9 consistent and 7 inconsistent, were then presented in a randomised order to each participant. The assignment of objects (circle, square, and triangle) was made at random to each set of assertions in the experiment.

The procedure was the same as in Experiment 1. Performance was self-paced, and the participants responded by pressing a key, labelled either "yes" or "no", to make their evaluation of whether each set of assertions could all be true at the same time.

Results

We excluded the data from 12 participants who responded faster than the criterion of a fast guess, namely, an evaluation of less than 7s; their mean latencies were 3.8s in comparison with a mean of 23.0s for the remaining participants. To reject so many participants is exceptional, and so we report additionally the results for all 32 participants below in the text. Table 4 presents the percentages of accurate evaluations for each of the 16 problems for the 20 remaining participants, and Figure 1 presents overall percentages and mean latencies for correct evaluations (in s). As the Figure illustrates, for the consistent sets the percentages of correct evaluations and their mean latencies were as follows:

Consistent 2. Mental model of the two conditionals satisfied the third assertion: 98% (20.2s).
Consistent 1. Mental model of one conditional satisfied the third assertion: 44% (22.9s).
Consistent 0. Mental model of no conditional satisfied the third assertion: 18% (27.7s).

The trends in accuracy (in percentage) and latency (in s) of correct evaluations were both significant (Page's $L = 274$, $z > 5.38$, $p < .001$, and $L = 129$, $z > 2.01$, $p < .05$, respectively). Because there is only one problem of the easiest sort, we also checked the difference between the consistent 1 and 0 problems for accuracy, which was also reliable (Wilcoxon test, $z > 2.84$, $p < .005$; $r = .45$). But, there were not enough correct evaluations in the consistent 0 condition for a test of the difference in latencies. The overall percentages of the correct

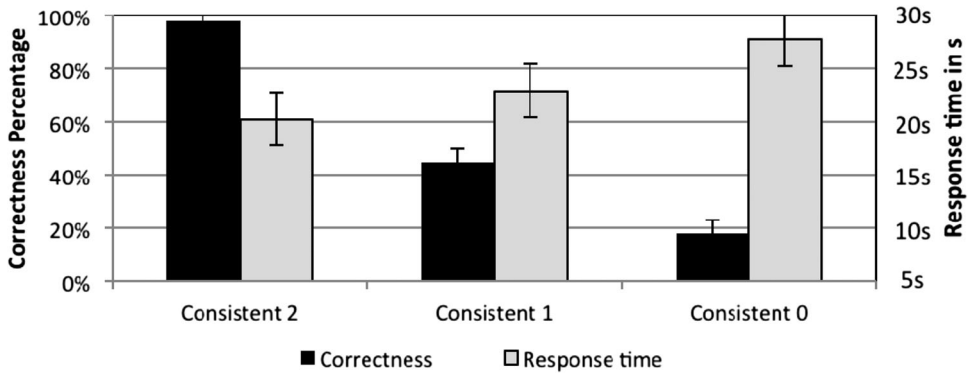

Figure 1. The percentages of correct evaluations and their latencies in s (with standard error bars) for the consistent problems in Experiment 4, depending on the number of conditionals with mental models satisfying the third assertion.

evaluations and their mean latencies for the two sorts of inconsistent problem were as follows

> Inconsistent 0. Mental model of no conditional satisfied the third assertion: 87% (22.1s).
> Inconsistent 1. Mental model of one conditional satisfied the third assertion: 65% (22.1s).

Only the difference in accuracy was significant (Wilcoxon tests, $z > 2.75$, $p < .006$; $r = .43$).

A reasonable worry about the preceding analyses is that they concerned only about two-thirds of the participants in the experiment. We therefore run subsidiary analyses of the results for all 32 participants. The pattern of accurate evaluations remained the same both for the consistent problems (consistent 2: 91% correct, consistent 1: 53% correct, and consistent 0: 28% correct (Page's L = 429.5, $z > 5.69$, $p < .00001$; and Wilcoxon test between consistent 1 and 0, $z > 3.57$, $p < .001$; $r = .45$), and for the inconsistent problems (inconsistent 0: 73% and inconsistent 1: 56%; Wilcoxon test, $z > 2.50$, $p < .01$; $r = .31$). Not surprisingly, the inclusion of the participants whose latencies seemed to reflect many guesses disrupted the reliable differences in the latencies of response: (consistent 2: 15.5s consistent 1: 17.9s; inconsistent 0: 16.0; Page's L = 383.0, $z > .12$, $p > .5$; inconsistent 1: 17.1s; inconsistent 0: 16.5 s; Wilcoxon test, $z = .44$, $p > .65$). Overall, however, the results corroborated the model theory's predictions for the evaluations of sets of assertions containing two conditionals.

General discussion

A basic conditional connecting spatial relations, such as:

> If the circle is to the left of the square then the square is to the left of the triangle.

has the mental models:

○ □ Δ
…

The first model represents the possibility in which both clauses in the conditional are true, and so it embodies a transitive relation: *the circle is to the left of the triangle*. The second model denoted with an ellipsis has no explicit content, and it represents the possibilities in which the *if*-clause of the conditional is false. System 1 with its limited computational power can construct mental models. It can also determine that each of the following assertions hold in the preceding explicit mental model:

> The triangle is to the right of the square.
> The square is to the right of the circle.

But, an assertion such as:

> The circle is to the right of the triangle.

does not match the explicit mental model. Individuals who access System 2 can build a fully explicit model in which both clauses of the conditional above are false:

Δ □ ○

This model in turn satisfies the preceding assertion, and a corresponding diagram. However, System 2's fleshing out of fully explicit models of complex conditionals puts a heavy demand on working memory, and so individuals should be more likely to rely on mental models in the present task, and therefore to succumb to the illusion that three assertions are inconsistent and that the diagram is not possible. This distinction between System 1 and System 2 is similar to other conceptions of dual processes in reasoning (e.g. Kahneman, 2011; but cf. Evans & Stanovich, 2013,

for an alternative account). System 1 underlies intuitions, and it is computationally weak, because it can run loops of operations for only a small finite number of times, and so it lacks even the power of a finite-state automaton (Hopcroft & Ullman, 1979). System 2 underlies deliberations, and it is computationally more powerful because it has access to working memory.

Our experiments showed that naive individuals can evaluate the consistency of sets of assertions containing complex conditionals, and that the model theory predicts their performance. When they could make the correct decision using System 1 and a mental model, the task was easy and they performed reliably better than chance. But, when they could make the correct decision only using System 2 and a fully explicit model, the task was difficult and they performed reliably worse than chance, succumbing to illusory inferences (Experiment 1). The same result occurred when the third assertion was replaced with question about the possibility of a spatial layout depicted in a diagram (Experiment 2). In addition, when the correct evaluation depended on coping with a second assertion in a set that described a transitive spatial relation, there was a small but reliable increase in the difficulty of consistent problems (see Table 1). When reasoners evaluated the consistency of conditionals and conjunctions, again the task was easier based on mental models than on fully explicit models (Experiment 3). The model theory predicts that in fleshing out fully explicit models of a conditional, *If A then B*, individuals tend to think first of the model, ¬*A and* ¬*B,* and then of the model, ¬*A and B*, where "¬" denotes an implicit negation. Although there was no overall difference in accuracy between these two conditions, the results did corroborate the trend in accuracy and the trend in the latency of evaluations, whether correct or not. A separate study of the consistency of simple conditionals bears out the same trend (Goodwin & Johnson-Laird, 2015). Only a small number of participants were able to engage System 2 and to cope systematically with evaluations that depended on fully explicit models.

Sets of assertions containing two conditionals allowed a more nuanced test of the model theory (Experiment 4). The correct evaluation of consistency was easiest when a third assertion held in a mental model common to both conditionals. It was more difficult when the assertion held in a mental model of only one conditional. And it was very difficult when the assertion held in no mental model of either conditional. The latter two sorts of evaluation depend on System 2, and again only a few participants were able to engage it in a reliable way. For inconsistent problems, the effect of mental models was again apparent. If the third assertion held in the mental model of one conditional, the participants were more likely to err than if it held in neither mental model of the conditionals. The first case created an illusion that the set is consistent, whereas in fact it was not.

Readers might suppose that reasoners can cope only with problems for which mental models suffice (using System 1), and that they can never cope with fully explicit models (using System 2). The idea is appealing, but our results refute it. If participants relied only on mental models, they would never make any correct evaluations that depended on fully explicit models. In fact, they make a small but reliable percentage of such evaluations. Could they merely be guessing? Sometimes perhaps, but the idea is not feasible in general. If participants only guessed, then their performance with inferences that depend on System 2 should be at chance, whereas it was in fact reliably lower than chance. Moreover, Experiment 4 provides a decisive rebuttal of this hypothesis. It cannot explain the difference between the two sorts of problem that depend on fully explicit models for their correct evaluation: consistent 1 problems (44% correct) and consistent 0 problems (22% correct). This difference, however, corroborates the model theory's prediction: participants try to use System 2, but, as the theory predicts, it is easier to construct a fully explicit model for one conditional (consistent 1 problems) than for two conditionals (consistent 0 problems).

The latencies in the experiments make sense. Apart from Experiment 2, all the experiments yielded mean latencies around 20–22s. Experiment 2 yielded very much faster latencies with a mean of about 11s. They were probably a consequence of the task: participants had only to determine whether a diagram depicted a possibility given two assertions, and did not have to evaluate whether a set of assertions was consistent. The experimental procedure split the sentences using the separate-stage paradigm of Potts and Scholz (1975), and so the first assertions disappeared on the presentation of the third assertion in Experiment 1 and of the diagram in Experiment 2. This procedure is closer everyday life when one has to encode initial assertions as they occur, and then to

try to integrate subsequent assertions within the encoding. Such a process is inimical to a purely verbal representation of the initial assertions, and so it may discourage verbal matching as a way to carry out the task (see Experiment 1). Although the procedure calls for holding an iconic representation in mind (Goodwin & Johnson-Laird, 2005; Johnson-Laird & Khemlani, 2013), it may also reduce the use of visual processes that impede reasoning (Knauff, 2013).

Is it necessary to postulate mental models in order to account for the results? There are at least three potential alternatives. The first alternative is that people rely on formal rules of inference to assess consistency. We described the general procedure for doing so in the Introduction, but it is highly implausible for individuals who have not mastered logic, because a correct evaluation of consistency depends on a failure to prove the negation of one assertion in a set from the remaining assertions. Such a failure, however, cannot account for the effect of the number of conditionals with mental models satisfying the third assertion in Experiment 4. A special case of this account is the "deductive" strategy, which we also outlined earlier: individuals judge that a set of assertions is consistent if, and only if, they can deduce one assertion from the others. But, this strategy fails as a general account, because assertions can be consistent when there are no deductive relations among them. Problem 2 in Experiment 1 is a good example:

If the square is to the right of the circle then the triangle is to the right of the square.

The circle is to the left of the triangle.
The square is to the left of the triangle.
Could all three of these assertions be true at the same time?

No assertion in the set can be proved to follow from the others. The second assertion follows from the conditional provided that its *if*-clause is true, but the third assertion is an implicit negation of the *if*-clause. Nevertheless, the explicit mental model of the conditional is:

○ □ Δ

Both the second and third assertions hold in this model, and the problem is quite easy (see Table 1). The strategy also fails to explain why evaluations of inconsistency can be rapid (as in Experiment 2): on its account, such evaluations ought to depend on an exhaustive attempt to find a proof that one assertion follows from the others, and so they should tend to take more time than evaluations of consistency. But, they do not. We conclude that individuals tend not to rely on the deductive strategy.

The second alternative is that individuals use a "suppositional" strategy (e.g. Evans, 2007). This account postulates that individuals assess their belief in a conditional using a revised version of Ramsey's (1929) test. To carry out the test, they add the conditional's *if*-clause hypothetically to their stock of knowledge, and then assess its *then*-clause. When they believe the *if*-clause, their evaluation of the *then*-clause establishes their belief in the conditional. If they do not believe the *if*-clause, Ramsey took the conditional to be void. But, in a revised version of the test, they modify their stock of knowledge to accommodate the false *if*-clause without inconsistency, and then assess the *then*-clause (Evans, 2007, p. 53). The theory has not as yet been applied to the assessment of the consistency of assertions, and it is not clear quite how it would apply. One apparent difficulty, however, is that the suppositional strategy leads to the so-called "defective" truth table for conditionals (Evans, 2007, p. 56 et seq.) in which a conditional, *If A then B*, is true in the case of *A and B*, false in the case of *A and not-B*, but has no truth value in the cases in which *A* is false. It follows, for example, that a set of assertions, such as:

If A then B.
¬B.
¬A.

cannot all be true at the same time. The reason is that the third assertion, ¬*A*, establishes that the *if*-clause of the conditional is false. The conditional itself is therefore neither true nor false (Evans, 2007, p.56), and so it cannot be true in any case in which the third assertion is true. Our participants begged to differ. Such evaluations are more difficult for them than those in which the *if*-clause of a conditional is true according to the other assertions, but their rate of judgments of consistency is well above zero, which is what the suppositional strategy seems to predict for such problems (see e.g. problems 3 and 4 in Table 1).

The third alternative is the "new paradigm" for reasoning, which is a confederation of theories that aim, in essence, to replace validity with probability (e.g. Adams, 1998; Evans, Handley, & Over, 2003; Oaksford & Chater, 2007; Pfeifer, 2013). They too tend to be based on Ramsey's test and the

defective truth table (for a review, see Johnson-Laird et al., 2015). Whether most reasoning is probabilistic is moot (cf. Johnson-Laird, 2006); much reasoning about uncertainty depends instead on possibilities. Likewise, the assessment of consistency is a major part of thinking in daily life: inconsistent beliefs, as we remarked at the outset, are a recipe for disaster. So, how might probabilities enter into assessments of consistency? Adams (1998) introduced the notion of probabilistic logic (p-logic). He argued that a major problem for logic is its treatment of conditionals, and so he proposed the defective truth table instead (see also de Finetti, 1937/1980). It follows from the defective truth table that the probability of a conditional, *If A then B*, equals the conditional probability of *B* given *A*. Adams argued that the meaning of a conditional *is* this conditional probability, so conditionals do not have truth values, and as a result they cannot enter into valid inferences. He therefore introduced the concept of probabilistic validity (p-validity) and an allied notion of p-consistency, which as far as we know has not been tested in any experiments. The essential idea is that if, and only if, each of the assertions in a set can have a high probability then they are p-consistent. For example, a pair of conditionals, such as:

If A then B.
If A then not B.

cannot both have high conditional probabilities: p(B|A) and p(¬ B|A). Hence, they are p-inconsistent, even though they are consistent in logic (Adams, 1998, Chap. 7).

Our results are contrary to p-consistency for three reasons. First, individuals judge that sets of assertions containing conditionals can all be *true*. According to p-logic, conditionals do not have truth values, and so the participants in our experiments should have baulked at the task. It should have been akin to asking whether a conditional that makes a request can be true, for example, "If you have some money, please lend me ten dollars." Requests really do not have truth values. So, according to p-consistency, our task is ill posed, and we should have asked participants instead, "Could all the assertions in the set have a high probability at the same time?"

Second, just as problems of the following sort (see problem 3 in Experiment 1):

If A then B.
¬B.
¬A.

are problematic for the suppositional strategy, so too they are problematic for p-consistency, because the third assertion refutes the *if*-clause of the conditional. There are various treatments of conditional probabilities in this case, ranging from treating the corresponding conditionals as certain (Adams, 1998, p. 181) to more recondite methods (see Cruz & Oberauer, 2014). On the first account, it should be easy to evaluate the set as p-consistent: the certainty of the conditional allows both the other assertions to have a high probability. In fact, the evaluation is difficult. The real problem for probabilistic approaches, however, is they offer no account of how individuals assess the probability of assertions, such as:

If the square is to the right of the circle then the triangle is to the right of the square.

Third, the problems that cause problems for the deductive strategy also yield data that p-consistency cannot explain. Consider, again, a problem of the sort (problem2 in Table 1):

If ′A then ′B.
C (a transitive conclusion from the mental model of the conditional).
B (an equivalent converse to ′B).

The set is straightforward for individuals to evaluate as consistent, but how does p-consistency explain this ease? Likewise, how can it account for the much greater difficulty of problems the following sort (problem 4 in Table 1)?

If A then B.
¬ C (a transitive conclusion from a fully explicit model of the conditional).
¬ A (an assertion that holds in the fully explicit model of the conditional).

The crux is that the new paradigm offers no explanation of how individuals infer that C (or ¬C) is consistent with the conditional. In our view, if the paradigm is to have any chance of explaining evaluations of consistency in these cases, it has to invoke inferential machinery that yields transitive inferences. Of course, not all relations are transitive, and, as the model theory predicts, some relations appear to be transitive but, in fact, are not (Goodwin & Johnson-Laird, 2008). Likewise, p-logic needs to explain both the three levels of difficulty in judgments of consistency and the two levels of difficulty in judgments of inconsistency in Experiment 4.

In conclusion, the theory of mental models explains how individuals evaluate the consistency

of assertions about conditional relations between spatial layouts. The process is straightforward when correct evaluations follow from System 1 operating with a single mental model. But, reasoners tend to rely on these models even when correct evaluations require System 2 to construct an alternative and fully explicit model. As a result, their inferences are illusory: they tend to judge that sets are inconsistent when, in fact, they are consistent, and, as our final study showed, to judge that sets are consistent when, in fact, they are inconsistent.

Acknowledgements

The authors are grateful to Sangeet Khemlani and Max Lotstein for their help, advice, and criticisms of a previous version of the paper. The authors thank three anonymous reviewers for stimulating criticisms.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG) with a Heisenberg-grant to the first author RA 1934/3-1 and the project R8-[CSPACE] as part of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition and in part by a grant to the third author from the National Science Foundation SES 0844851 to study deductive and probabilistic reasoning.

References

Adams, E. W. (1998). *A primer of probability logic*, Stanford, CA: CSLI Publications.

Barrouillet, P., Grosset, N., & Lecas, J. F. (2000). Conditional reasoning by mental models: Chronometric and developmental evidence. *Cognition*, *75*(3), 237–266.

Barrouillet, P., & Lecas, J. F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *5*, 289–302. doi:10.1080/135467899393940

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

Byrne, R. M., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory & Language*, *28*, 564–575. doi:10.1016/0749-596X(89)90013-2

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.

Cruz, N., & Oberauer, K. (2014). Comparing the meanings of "if" and "all". *Memory & Cognition*, *42*, 1345–1356. doi:10.3758/s13421-014-0442-x

Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. St. B. T., Handley, S. J., & Over, D.E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–355. doi:10.1037/0278-7393.29.2.321

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives in Psychological Science*, *8*, 223–241. doi:10.1177/1745691612460685

de Finetti, B. (1937/1980). Foresight: Its logical laws, its subjective sources (H. E. Kyburg, English trans.). In J. H. E. Kyburg & H. E. Smokier (Eds.), *Studies in subjective probability* (pp. 55–118). New York: Robert E. Krieger.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*, 468–493. doi:10.1037/0033-295X.112.2.468

Goodwin, G., & Johnson-Laird, P. N. (2006). Reasoning about the relations between relations. *Quarterly Journal of Experimental Psychology*, *59*, 1047–1069. doi:10.1080/02724980543000169

Goodwin, G. P., & Johnson-Laird, P. N. (2008). Transitive and pseudo-transitive inferences. *Cognition*, *108*, 320–352. doi:10.1016/j.cognition.2008.02.010

Goodwin, G. P., & Johnson-Laird, P. N. (2015). The truth of conditional assertions (Unpublished).

Hopcroft, J. E., & Ullman, J. D. (1979). *Formal languages and their relation to automata*. Reading, MA: Addison-Wesley.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*, 2075–2087. doi:10.3758/BF03192939

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

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2006). *How we reason*. New York: Oxford University Press.

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. doi:10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., Khemlani, S. S. (2013). Chapter one - toward a unified theory of reasoning. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 59, pp. 1-42). Urbana, IL: Academic Press. doi:10.1016/B978-0-12-407187-2.00001-0

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*, 640–661. doi:10.1037/0033-295X.111.3.640

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning.

Trends in Cognitive Sciences, *19*, 201–214. doi:10.1016/j.tics.2015.02.006

Johnson-Laird, P. N., Legrenzi, P., Girotto, P., & Legrenzi, M.S. (2000). Illusions in reasoning about consistency. *Science*, *288*, 531–532. doi:10.1126/science.288.5465.531

Juhos, C., Quelhas, C., & Johnson-Laird, P. N. (2012). Temporal and spatial relations in sentential reasoning. *Cognition*, *122*, 393–404. doi:10.1016/j.cognition.2011.11.007

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.

Khemlani, S., & Johnson-Laird, P.N. (2013). The processes of inference. *Argument and Computation*, *4*, 4–20. doi:10.1080/19462166.2012.674060

Khemlani, S., Lotstein, M., & Johnson-Laird, P. N. (2015). Naive probability: Model-based estimates of unique events. *Cognitive Science*, *39*, 1216–1258. doi:10.1111/cogs.12193

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2013). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*, 541–559. doi:10.1080/20445911.2012.660913

Khemlani, S. S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P.N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Sciences*, *110*, 16766–16771.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. Cambridge: The MIT Press.

Knauff, M., & Ragni, M. (2011). Cross-cultural preferences in spatial reasoning. *Journal of Cognition and Culture*, *11*, 1–21. doi:10.1163/156853711X568662

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York: Oxford University Press.

Pfeifer, N. (2013). The new psychology of reasoning: A mental probability logical perspective. *Thinking & Reasoning*, *19*, 329–345.

Potts, G. R., & Scholz, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning and Verbal Behavior*, *14*, 439–452. doi:10.1016/S0022-5371(75)80023-5

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*, 561–588. doi:10.1037/a0032460

Ragni, M., Knauff, M., & Nebel, B. (2005). A computational model for spatial reasoning with mental models. In B. Bara, B. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1064–1070). Mahwah, NJ: Erlbaum

Ramsey, F.P. (1929/1990). General propositions in causality. In D. H. Mellor (Ed.), *Philosophical papers* (pp. 145–163). Cambridge: Cambridge University Press.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge: The MIT Press.

Schaeken, W. S., Girotto, V., & Johnson-Laird, P. N. (1998). The effect of an irrelevant premise on temporal and spatial reasoning. *Kognitionswissenschaft*, *7*, 27–32. doi:10.1007/BF03354960

Schaeken, W. S., Johnson-Laird, P. N., & d'Ydewalle, G. (1996a). Mental models and temporal reasoning. *Cognition*, *60*, 205–234. doi:10.1016/0010-0277(96)00708-1

Schaeken, W. S., Johnson-Laird, P. N., & d'Ydewalle, G. (1996b). Tense, aspect, and temporal reasoning. *Thinking & Reasoning*, *2*, 309–327. doi:10.1080/135467896394456

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Wason, P. C., & Johnson-Laird, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, *61*, 509–515. doi:10.1111/j.2044-8295.1970.tb01270.x