# Reasoning about Spatial Consistency

**Marco Ragni (ragni@cs.uni-freiburg.de)**
Department of Psychology, University of Giessen, Giessen, Germany
Department Foundations of AI, Technical Faculty, University of Freiburg, Freiburg, Germany

**Tobias Sonntag (sonntag@hotmail.de)**
University of Freiburg, Freiburg, Germany

**P.N. Johnson-Laird (phil@princeton.edu)**
Department of Psychology, Princeton University, NJ, US
Department of Psychology, New York University, USA

**Abstract**

The consistency of spatial descriptions is relevant to tasks ranging from navigation to architecture. In contrast to studies of deduction in which a conclusion is drawn from premises, there have been only a few investigations into how human reasoners decide whether or not a description is consistent. We report results corroborating the theory that reasoners make such judgments usually relying on a single initial mental model of the description. As a result, the task is difficult if it calls for an alternative model of the assertions that must be revised. Especially the model construction process and the way of how information is integrated into a model can explain errors in evaluating problems as consistent. Implications for other theories of reasoning are discussed.

**Keywords:** Consistency; Spatial Relational Reasoning; Mental Model Theory

## Introduction

Inconsistency in a set of beliefs or assertions is dangerous, and can have disastrous consequences. Its importance therefore raises the question of how individuals assess consistency – that is, the assertions or beliefs can all hold at the same time. We have investigated this problem using descriptions of spatial layouts, which have everyday analogs in architecture, route finding, and design. Consider, for example, the following problem about a fruit and veg stall:

(1) The box of apples is left of the box of pears.
The box of kiwis is right of the box of the pears.
The box of apples is right of the box of kiwis.

Can these three assertions all be true at the same time?

The answer to this question is "No", as there is no arrangement of the three boxes in a line integrating all information in the description. We therefore refer to a set of assertions as *consistent* if there it has at least one model sattisfying all the assertions in the set. A general test for inconsistency based on formal logic works as follows: Choose any assertion from the set of assertions, and prove that its negation follows from the remaining assertions in the set. Only if no such proof exists is the set consistent. Hence, consistency depends on the failure of an exhaustive search for all possible proofs. – a process that is computationally intractable. An alternative process could be based on the new paradigm of probabilistic logic. Adams (1998) formulated a notion of p-consistency, according to which a set of assertions is consistent if each assertion in the set can have a high probability. No psychologists, as far as we know, have endorsed this procedure. One difficulty is to specify how people determine the relative constraints on the probabilities of assertions in a set. So, we need an alternative account. In the next section, we therefore describe the mental model theory – the "model" theory for short – and we derive its predictions for assessments of consistency. We then report an experiment that tested these predictions. Finally, we discuss the implications of these results.

## The mental model theory of consistency

Consider the following problem:

(2) The apple is to the left of the pear.
The pear is to the left of the kiwi.
The pear is to the left of the orange.
The orange is to the left of the mango.
The kiwi is to the left of the orange.

Can these five assertions all be true at the same time?

As you read these assertions, you can construct a model of the corresponding spatial arrangement:

apple pear orange mango kiwi

You may have formed a visual image of the arrangment, or your representation may have been more abstract. It needs only to represent the spatial relations among the objects (Goodwin & Johnson-Laird, 2005; Knauff 2013). You may have noticed that the *orange* can initially be located either to the left or to the right of the *kiwi,* but the final assertion resolves the interminancy. The example illustevaluates a *a* temporary spatial indeterminacy (e.g., Johnson-Laird & Byrne, 1991): although the set of assertions yields a determinate arrangement, during their interpretation more than one arrangement is possible. Likewise, reasoners often initially construct a *preferred* mental model, and neglect other possible mental models (e.g., Ragni & Knauff, 2013).

This leads us to the first research question: *Although all assertions form a determinate description has the indeterminacy during the construction process an influence on reasoning performance?* If so, this would not only support a model based approach, but show that the model construction process is a relevant factor in deciding consistency. Preferred models are incrementally constructed, i.e., during the construction process each new premise information is taken incrementally into account. Such a model construction process saves working memory capacity, since each bit of information is immediately processed and integrated into the model (Johnson-Laird & Byrne, 1991). For the different construction principles please refer to Table 1.

**Table 1.** Construction principles for the first four assertions leading to three possible models in the indeterminate problem (2). The fifth assertion eliminates all models but the *right-most insertion model*. The asterisk denotes the initial model.

| Insertion principle | Model |
|---|---|
| Right-most insertion* | apple pear kiwi **orange mango** |
| Mix-Left/Right insertion | apple pear **orange** kiwi **mango** |
| Left-most insertion | apple pear **orange mango** kiwi |

Consequently, participants may construct the following mental model for the first four assertions of Problem 2:

apple pear kiwi orange mango
. . .

where the ellipsis denotes implicit models – in this case the other models that can be found in Table 1. The fifth assertion is consistent with these possibilities because it holds in the explicit mental models, and the model that is present to the participants is the same as the explicit mental models (all other models in Table 1), and yields the response: "yes (all assertions are consistent)." In contrast, consider the following problem:

(3) The apple is to the left of the pear.
The kiwi is to the right of the pear.
The orange is to the right of the pear.
The orange is to the left of the mango.
The mango is to the left of the kiwi.

Again, can these five assertions all be true at the same time? This set of five assertions is not consistent with the initial model (apple pear kiwi orange mango). The fifth assertion 'The mango is to the left of the kiwi' forces the reasoner to revise the recently constructed initial model to the model

apple pear orange mango kiwi

as the fifth assertion conflicts the previous built model. The fifth assertion (in problem 3) does not hold in the preferred or initially constructed model (built after the assertions 1-4, cp. Table 1), but only in the model constructed according to the leftmost insertion principle. In this sense we would have a mismatch between the initial model (the model: apple pear kiwi orange mango) and the fifth assertion ("The mango is to the left of the kiwi"). Of course, all five statements are consistent.

Thus, the theory of mental models predicts that reasoners may have some difficulty in inferring that problems such as 3 are consistent, as the participants will have a conflict with their initial model they constructed after the first four assertions. The fifth assertion does not correspond with the initial model of the four assertions and so individuals should respond, "no" – in contrast to Problem 2.

Human reasoners tend to evaluate a given set of assertions as inconsistent if it does not match the initially built mental model, which is constructed according to the *right-most insertion principle* (see Table 1) – that is an alternative name for the first-free-fit principle (Ragni & Knauff, 2013). This initial mental model is the central explanation pattern in reasoning towards consistency. Even if a description is consistent, a failure to build this model will result in an erroneous answer.

If participants construct initial models then the way information is integrated should have as well an influence. This effect has been investigated for deductive reasoning in the *premise order effect*. Knauff, Rauh, Schlieder and Strube (1998) conducted an experiment to test the empirical differences of continuous, semi-continuous, and discontinuous premise orders in spatial relational reasoning (following Ehrlich & Johnson-Laird, 1982). The continuous order and the semi-continuous order led to 60% correctness and the discontinuous order to 50% only. The premise order effect is explained with the effort to construct a mental representation of the premises. In continuous and semi-continuous descriptions, a common middle term of two successive premises exists. Since this is not the case in discontinuous premise orders, these premises are more difficult to process and we will leave these problems out. Again, if participants are successively integrating information than reasoners struggle more when drawing valid conclusions from a set of assertions that cannot be successively integrated into one initial model. In the continuous premise order condition (cp. Table 2), each assertion (but the first) contains one new introduced object.

Both aspects – the way and kind of the construction of the initial mental model are the two main predictions of the mental model theory to explain human evaluation of consistency and will be investigated in the following.

## The experiment

The participants' task was to evaluate whether or not spatial descriptions were consistent. Half of the problems were determinate and half of them had a local inderminacy that the fifth assertion resolved. One third of the indeterminate problems asked for the preferred model (no mental model revisions necessary), one third asked for the alternative models with a revision distance 1 from the preferred model and one third asked for the alternative models with a revision distance 2 from the preferred model.

**Table 2.** Determinate (left column) and indeterminate (right column) one-dimensional problems. We obtained a semi-continuous premise order by exchanging the first two premises, which implies a directional change during the postulated model building rather than the fleshing out of a sub-model. All of these problems were also available in an inconsistent and vertical version.

| | **Determinate problems** | **Indeterminate problems** |
|---|---|---|
| premise number | continuous premise order | |
| 1 | The apple left of the pear | 1 The apple left of the pear |
| 2 | The pear left of the kiwi | 2 The pear left of the kiwi |
| 3 | The kiwi left of the orange | 3 The pear left of the orange |
| 4 | The orange left of the mango | 4 The orange left of the mango |
| 5 | The pear left of the mango | 5 The pear left of the mango |
| | semi-continuous premise order | |
| 2 | The pear left of the kiwi | 2 The pear left of the kiwi |
| 1 | The apple left of the pear | 1 The apple left of the pear |
| 3 | The kiwi left of the orange | 3 The pear left of the orange |
| 4 | The orange left of the mango | 4 The orange left of the mango |
| 5 | The pear left of the mango | 5 The pear left of the mango |

### Participants

We tested 27 logically naive participants (15f/12m; mean age 27 years) on an online website, Amazon's Mechanical Turk, and paid them a nominal fee for their participation.

### Materials

The problems consisted of five assertions stating spatial relations of five objects. Each of the first four successive assertions introduced a new object, which was randomly inserted from a list of either fruits (*apple*, *peach*, *orange*, etc.) or "breakfast items" (*toast*, *bagel*, *biscuit*, etc.). The forty problems differed in five independent variables with respect to *consistency, determinacy of the description, premise order, distance, and type of relation.*

Half of the problems were consistent (e.g., Problem 2) and half of the problems inconsistent (e.g., Problem 1). We also manipulated the *determinacy of the description,* so that half of the problems' descriptions are determinate (see, Table 2), i.e., they allow for only one model after four presented assertions, and half of them were indeterminate, i.e., they allowed for three possible arrangements of the objects. This indeterminacy appears, however, only in the first four assertions, after the fifth assertion, each problem description is determinate. The indeterminate problems differed variable in the *revision distance of the initial model* to integrate the last assertion. It can require zero vs. one vs. two spatial operations from the initially built mental (see Table 1).

The third variable is the *sequence*, i.e., are the first two assertions continuous or semi continuous (see Table 2). We manipulated this by exchanging the first two assertions. We counterbalanced the problems regarding the *type of relation*, i.e., half of the problems used horizontal relations like left and right and half of the problems vertical relations (above and under).

### Design and Procedure

Our 40 problems differed in the three variables outlined above: Determinate vs. indeterminate problems, consistent vs. inconsistent, and the sequence of the premises. In order to examine possible effects of revising initial models, we only considered the consistent indeterminate tasks and differentiated the type of conflict to the initial model. First, only for consistent tasks we can interpret a negative response as an indicator of difficulty. Second, for these problems a conflict to the initial model does not mean a conflict to its consistency. In the condition '0' (the *rightmost insertion principle* in Table 1), the representation of an initial model was possible till the last introduced assertion. In 1-step-revised-model condition (the *mix-left/right-insertion* principle in Table 1), a revision of the model only required the revision of two objects of the initial model. In the 2-revised-model condition (the *leftmost-insertion principle* in Table 1), more operations were needed to revise the initial model and detect the problems' consistency. Assuming the representation of an initial model, we expected increasing difficulty related to the number of operations necessary.

Each participant received all 40 problems and acted as their own controls. Each problem and each assertion were presented self-paced. After the participants received the fifth assertion, which either was consistent or inconsistent with the four preceding assertions (consistent vs. inconsistent), they had to consider the consistency of each problem. Participants could verify ("*y*") or reject ("*n*") all the assertions by answering the question "*Could all of these assertions be true at the same time?*" The problems were presented in a randomized order. Reading and response times were recorded as well as the correctness of answers and analyzed as dependent variables.

### Results and Discussion

Six participants were excluded from the analysis, as their accuracy did not differ significantly from chance. The remaining 21 participants solved an average of 85% of all problems correctly. Table 3 presents the percentages of correct response in each of the 4 main conditions.

Reasoners correctly identified as consistent determinate descriptions (90%) more often than descriptions that were locally indeterminate (72%, Wilcoxon test, $z = 3.08$, $p < .01$ [1-tailed]) and the same pattern holds for the reaction times (29.0 vs. 23.4, Wilcoxon test, $z = 2.88$, $p < .01$ [1-tailed], see Figure 1). The participants identified inconsistent problems (89%) in general significantly more often than consistent ones (82%, $z = 2.12$, $p < .05$ [2-tailed], r = -.46).

**Table 3.** Aspects of the experimental problems, correctness in percentage and response times in seconds. Half of the problems are determinate and in half of the problems the first four assertions were indeterminate and allowed for three models (initial model, 1-step revised initial model, and 2-step revised initial model). .

| Problems | Correctness in % | Response times |
|---|---|---|
| Overall mean | 85 | 26.8s |
| Determinate descriptions | 90 | 29.0s |
| Indeterminate descriptions | 72 | 23.4s |
| Models after 4th premise: | | |
| *Initial model* | 93 | 24.3s |
| *1-step revised initial model* | 80 | 30.7s |
| *2-step revised initial model* | 49 | 31.3s |
| Continuous assertions | 85 | 25.1s |
| Semi-continuous assertions | 84 | 28.4s |

In the indeterminate condition the first four assertions allowed for three different models (cp. Table 1), while with the fifth assertion the set of assertions was again a determinate description and, hence, allowed for one model only. The more transformation steps the fifth assertion requires from the initial model the lower was the correctness rates and the higher are the response times: If there is no revision step necessary (i.e., if the fifth assertion is consistent with the initial model the correctness is 93%); if the fifth assertion requires one revision-step of the initial model (cp. Table 1) the accuracy and response times are lower than in the initial model condition (80%, Wilcoxon test, $z = 1.98$, $p < .01$; response times: 24.3s vs. 30.7s, Wilcoxon test, $z = 1.68$, $p < .05$). The same pattern holds if two model revisions are necessary (correctness 49%, Wilcoxon test, $z = 3.36$, $p < .001$; response times: 24.3s vs. 31.3s, Wilcoxon test, $z = 1.90$, $p < .05$). The manipulation of the first two assertions' order (the continuous order vs. semi-continuous order cp. Table 2) did not effect mean correctness to a significant extent (85% vs. 84%, Wilcoxon test, $z = .34$, $p = .735$). However, participants needed significantly more time to generate answers, in the semi-continuous case (when the first two assertions were exchanged; 25.1s vs. 28.4s, Wilcoxon test, $z = 2.17$, $p < .05$ [1-tailed], $r = -.46$). This delay can be traced back to longer reading times for the third assertion in the semi-continuous order in contrast to the continuous order. This supports again a continuous integration of information into a model during the reasoning process. In accordance with previous results (Ragni, Fangmeier & Schleipen, 2007) orientation of relations differed between the postulated vertical model building (87%) prompted by the relation *above*, and the supposed horizontal model building prompted by the horizontal relation *left* (83%, Wilcoxon test, $z = 1.77$, $p < .05$ [1-tailed]).

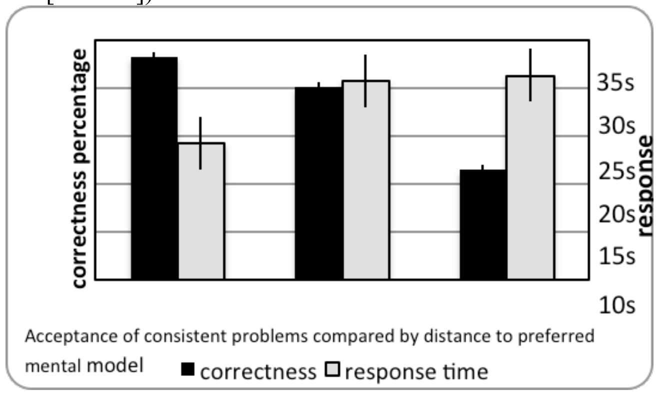


**Figure 1.** Correctness rates (left abscissa, black bar) and response times (right abscissa, grey bar) for consistent problems compared for the fifth assertion true in the initial-model condition (0), 1-step revised-model condition (1), and 2-step revised-model condition (2) with ascending transformation distance to the initial model implicated by the first four assertions.

## General Discussion

In contrast to an evaluation in formal logic, the order of assertions has a major effect on human evaluations of consistency. This finding occurs even when a description is determinate and consistent, e.g.:

The apple is to the left of the pear.
The pear is to the left of the kiwi.
The pear is to the left of the orange.
The orange is to the left of the mango.
The kiwi is to the left of the orange.

Can all five of these assertions be true at the same time?

Reasoners can construct a model of the first two assertions:

apple pear kiwi

But, how are they to interpret the third assertion? The orange could be to the right of the kiwi or it could be between the pear and the kiwi. The final assertion in the description resolves the indeterminacy, but nevertheless its local occurrence impedes the evaluation of the description as consistent. Indeed, if a subsequent assertion contradicts an initial model, then the chances increase that reasoners will err and evaluate the description as inconsistent.

Similar difficulties occur when the referents in a description are ordered discontinuously, e.g.:

> The apple is to the left of the pear.
> The orange is to the left of the mango.
> The pear is to the left of the orange,
> etc.

The second assertion cannot be integrated into the model of the first assertion until reasoners interpret the third assertion. This discontinuity contributes to the difficulty of evaluating consistency. But, theories that do not postulate the construction of mental models have difficulty in explaining the phenomenon.

A byproduct of our investigation was the finding that human reasoners find it slightly easier to work in a vertical direction, e.g., *A is above B*, than in a horizontal direction, e.g., *A is to the left of B* (see Ragni, Fangmeier, & Schleipen, 2007, for similar results). Why the difference occurs remains an open question, but studies of spatial orientation have shown that individuals are less likely to confuse vertical relations with left-to-right relations (Sholl & Egeth, 1981).

The two main alternatives to the model theory – mental logic and probability logic – have not addressed reasoning about the consistency of spatial descriptions. The only general method for testing consistency in mental logic is to negate one assertion, and to try to prove that it follows from the remaining assertions (e.g., Rips, 1994). If it does, then the description is inconsistent; if it doesn't then the description is consistent provided that one has made an exhaustive search and the logic is complete. The case of consistency can lead to a potentially exponential blow-up of the applications of rules governing the transitivity of spatial relations. In contrast, an inconsistency can be discovered in a single proof that the negated assertion follows from the remaining assertions. The notion that naïve reasoners grasp these principles seems unlikely. Moreover, the account fails to explain the difference in difficulty between two consistent problems: problem 2 was evaluated correctly on 93% of trials, whereas problem 3 was evaluated correctly on only 49% of trials. The model theory predicts the difference because problem 2 is straightforward whereas problem 3 has a fifth assertion calling for reasoners to revise their model of the earlier assertions. Neither problem yields a proof than the negation of one assertion follows from the other assertions.

The difference in difficulty between problems in a vertical dimension and those in a horizontal dimension makes sense in the model theory: a confusion between left and right should be echoed in the construction of models. But, it is inexplicable for theories based on formal rules: there is no reason why the transitivity of *above* should be easier to grasp than the transitivity of *left*. The concept of lexical marking according to which marked items are harder to work with than unmarked items (see, e.g., Clark, 1969; Evans, Newstead, & Byrne, 1993). Both *left* and *right* are marked, whereas *above* is unmarked and *below* is marked. The difference might potentially account for our result.

Taken together the mental logic theory leaves the questions about the reasoning performance differences in indeterminate cases open. For deductive reasoning Van der Henst (2002) proposed to extend the set of reasoning rules for rules of indeterminacy. This would, however, not help in our case than all five assertions together are a determinate description of the problems and thus not requiring any mental logic rules of indeterminacy.

Probabilistic approaches can explain deductions from conditional and quantified premises (e.g., Oaksford & Chater, 2001). But, the evaluation of consistency challenges this approach (Kunze *et al*. (2011). As we have argued in the introduction, psychologists have not applied the notion of probabilistic-consistency (Adams, 1998) to human reasoning. As in the case of theories based on logic, it is not obvious how it can explain our principal results. A further difficulty is to account for how people estimate the relative probabilities of spatial assertions.

In contrast, to its alternatives, the model theory provides a simple explanation of the phenomena. If, and only if, individuals can build a model of a set of assertions then they judge them to be consistent. An initial model may clash with a subsequent assertion. Reasoners may then search for an alternative model to accommodate the assertion. Even if they find one, the task is harder than when the initial model accommodates all the subsequent assertions. Likewise, the task will be harder when there is a discontinuity in the referents. Reasoners have to bear in mind two separate spatial relations, which they can integrate only in the light of a subsequent assertion. Once again, this factor adds to the difficulty of evaluating consistency.

In conclusion, reasoning about the consistency of descriptions is important in everyday life. The model theory provides an account of how naïve reasoners carry out this task, and our investigation has corroborated its main predictions.

## Acknowledgements

The research was partially supported by the DFG with a Heisenberg grant (RA 1934/3-1) and a DFG research grant (RA 1934/2-1) in the SPP 1516 "New frameworks of Rationality" to the first author. The authors are grateful to Stephanie Schwenke for proof-reading and Eva-Maria Steinlein for comments and discussions.

## References


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

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

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, *76*, 387-404.

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior,* 21, 296-306.

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

Goodwin, G. P. & Johnson-Laird, P. N. (2005). Reasoning about relations, *Psychological Review*, 112, 468-493.

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

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

Knauff, M. (2013). *Space to reason*. Cambridge, MA: MIT Press.

Knauff, M., Rauh, R., Schlieder, C. & Strube, G. (1998). Continuity effect and figural bias in spatial relational inference. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 573-578.

Kunze, N., Khemlani, S., Lotstein, M. & Johnson-Laird, P. (2010). Illusions of Consistency in Quantified Assertions. In Proceedings of the 31th Cognitive Science Conference (573–578). Mahwah: Erlbaum.

Oaksford, M. & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, 5, 349-357.

Ragni, M., Fangmeier, T., & Schleipen, S. (2007). What about negation in spatial reasoning? *Proceedings of the 28th Annual Cognitive Science Conference*, 1409–1414.

Ragni, M., Fangmeier, T., Webber, L. & Knauff, M. (2007). Preferred mental models: How and why they are so important in human reasoning with spatial relations. *Spatial Cognition V: Reasoning, Action, Interaction*, 175-190.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, 120 (3), 561–588.

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

Sholl, M. J., & Egeth, H. E. (1981). Right-left confusion in the adult: A verbal labeling effect. *Memory & Cognition*, *9*, 339-350.

Van der Henst, J. (2002). Mental model theory versus the inference rule approach in relational reasoning. *Thinking & Reasoning*, 8, 193-203.