Illusions of consistency in quantified assertions

Niklas Kunze[1] (niklas.kunze@uni-konstanz.de)
Sangeet Khemlani[2] (khemlani@princeton.edu)
Max Lotstein[2] (lotstein@princeton.edu)
P.N. Johnson-Laird[2] (phil@princeton.edu)
[1]Department of Psychology, University of Konstanz, 78457 Konstanz, Germany
[2]Department of Psychology, Princeton University, Princeton, NJ 08540, USA

Abstract

The mental model theory of reasoning postulates that individuals establish the consistency of a set of assertions by constructing a mental model in which all the assertions hold. Mental models represent what is true but not what is false, and this principle of 'truth' predicts that certain assertions should yield systematic errors. We report an experiment in which participants evaluated the consistency of assertions based on quantifiers and sentential connectives, e.g., *All of the artists are barbers or else all of the barbers are artists; Some of the artists are not barbers*. The results showed that participants judged consistent assertions to be inconsistent, and vice versa, much more often for the predicted assertions than for control problems, which should be unaffected by the failure to represent what is false. These results provide a litmus test for mental models, because no current alternative theories of reasoning predict them.

Keywords: deductive reasoning; mental models; consistency; illusions.

Introduction

Are humans inherently rational? Without any training, they are able to make valid deductions. Consider the following problem:

> Carol invested in capital securities or else she invested in municipal bonds.
> She did not invest in municipal bonds.
> Therefore, she invested in capital securities.

You do not need to know anything about securities or bonds to tell that the inference is valid. The ability to make valid inferences is a cornerstone of rationality, and as such, many theories argue that humans make use of formal rules akin to those in logic (Braine & O'Brien, 1998; Rips, 1994). According to those theories, humans make mistakes because they misapply the rules. Likewise, theories based on a probabilistic calculus believe that humans are rational, and that cognitive scientists use the wrong criteria to assess rationality, because everyday reasoning is probabilistic (Oaksford & Chater, 1998, 2007). Our own alternative theory is that human reasoning is based on *mental models*, or iconic representations of possibilities (Johnson-Laird, 2006; Johnson-Laird & Byrne, 1991). Mental models represent what is true and not what is false, and this constraint can lead to inaccurate models of assertions. Thus, human reasoning is fallible in practice, and individuals should succumb to systematic errors in judgment and inference. These errors are at present a unique prediction of the model theory, and so they serve as a litmus test for mental models. In the present paper we examine fallacious judgments of consistency for assertions combining quantifiers such as 'all' and connectives such as 'or else'.

Mental models and illusions

A foundational assumption of the model theory is the *principle of truth*: the mental models of a set of assertions represent only those possibilities consistent with the truth of assertions. The principle applies to assertions as a whole as well as to clauses within them. For example, an exclusive disjunction *A or else not B* yields the following mental models, where each horizontal line represents a model of a possibility, and '¬' is used to denote negation:

A	
	¬ B

The models do not represent what is false according to the disjunction, such as the case in which *A* is false and *B* is true. And, for those possibilities that make the exclusive disjunction true, such as the case in which *A* is true, the falsity of the corresponding possibility, in this case the negation of ¬*B*, hence *B*, is not represented in the models. This means that a *literal* (a proposition such as ¬*B* that contains no sentential connective) is represented in a model only if it is true in a possibility. Thus, the first of the models above represents the possibility that A is true, but it does not represent the fact that the literal ¬*B* is false in the possibility. Mental models do not represent what is false, whether it is an affirmative or negative literal, but in certain cases, such as when an inferential task is easy, individuals can construct *fully explicit models*. They represent both what is true and what is false in each possibility and therefore yield the correct representation of the assertion. The fully explicit models of *A or else not B* are as follows:

A	B
¬ A	¬ B

where the affirmative *B* in the first model represents the falsity of the negation, ¬*B*, and the negative ¬*A* in the second model represents the falsity of affirmative, *A*. As these models show, the disjunction is equivalent to the biconditional: *A if and only if B*. However, very few people grasp the equivalence, because it is evident only to those who envisage what is false and construct fully explicit models.

The principle of truth may seem like a useful compromise to reduce the load on working memory, but it has an unexpected consequence: it predicts *illusions*, i.e., judgments and inferences that are compelling but erroneous (Johnson-Laird & Savary, 1999). Illusions can occur when individuals assess conclusions, make inferences, or evaluate the consistency of a set of assertions. For example, consider this problem:

> There is a pin and/or a bolt on the table, or else a bolt and a nail on the table.
> There is a bolt and a nail on the table.
> Is it possible that both assertions could be true at the same time?

Reasoners in one study overwhelmingly responded 'yes' (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000), but the response is a fallacy predicted by the principle of truth. It is a fallacy because 'or else' in the first premise is an exclusive disjunction, i.e., if one clause of the disjunction is true, the other must be false. Hence, the truth of the second premise, *there is a bolt and a nail on the table*, implies that both clauses in the first premise are true. And that contravenes the meaning of 'or else', which means that the two assertions cannot be true at the same time. However, reasoners do not grasp this inconsistency and instead incorrectly judge the two assertions to be consistent.

Previous studies have corroborated the existence of illusory *deductions* from disjunctive and biconditional premises (Johnson-Laird & Savary, 1999). They have also corroborated them in singly quantified premises (Yang & Johnson-Laird, 2000). But, no study has examined the interaction between connectives and quantifiers. Our aim was accordingly to test whether illusions also occurred in a new domain: the evaluation of the *consistency* of assertions that depend on both quantifiers and connectives (as in Problem 1 below).

Illusions with quantified assertions

Our experiment examined two sorts of assertions that should yield illusions: exclusive disjunctions of quantified assertions, such as *All the A are B or else some of the B are A,* and biconditionals of quantified assertions, such as *All of the A are B if and only if All of the B are A*. Half of the problems used in the study were those that the principle of truth predicts should yield illusory judgments of consistency, and the other half were those that the principle of truth predicts should yield correct responses. Here is an example of an illusory problem based on an exclusive disjunction:

> *1. Illusion (disjunction)*
> All of the artists are barbers or else some of the barbers are artists.
> All of the barbers are artists.
> Is it possible for both statements to be true at the same time?

The disjunction yields two mental models that represent the two clauses (*All of the artists are barbers* and *Some of the barbers are artists*). Each model contains a set of individuals, where each line represents an individual and denotes the individual's properties. We lay out the two models as follows:

1.		2.	
[artist]	barber	barber	artist
[artist]	barber	barber	artist
		barber	

These models represent each set by a small but arbitrary number of individuals (two or three in the present models), and the square brackets denote that a set has been represented exhaustively (cf. the notion of 'distribution' in logic). One consequence of these exhaustively represented properties is that they cannot be added to new individuals in the model (see, e.g., Johnson-Laird, 2006). So, you cannot add instances of artists that are not barbers to the first model.

Consider the first mental model, which represents the first clause of the disjunction, *All of the artists are barbers*. The second assertion in the problem, *All of the barbers are artists*, is true in this model. The model theory predicts that individuals judge a set of assertions to be consistent if all the assertions hold in at least one mental model. Hence, people should judge that the two assertions are consistent. However, this judgment is flawed. The principle of truth predicts that the mental models represent the truth of each clause in an exclusive disjunction, but not the concurrent falsity of the other clause. Suppose that the first clause in the disjunction, *All of the artists are barbers*, is true. Hence, the second clause must be false, i.e., none of the barbers is an artist. This case is inconsistent with the first clause of the disjunction, and so it is impossible. Now suppose that the second clause of the disjunction is true, i.e., some of the barbers are artists. In this case, it must be false that all of the artists are barbers, i.e., at least some of them are not barbers. So, we have the conjunction of at least some of the barbers are artists and at least some of the artists are not barbers. There is accordingly just one fully explicit model of the disjunction:

artist	barber
artist	barber
	barber
artist	¬barber

The second assertion in the problem, *all the barbers are artists*, is accordingly inconsistent with this model, and the correct evaluation of the two assertions is that they are inconsistent.

The same compound assertion used in (1) yields a control problem, as in (1'):

1'. Control (disjunction)

All of the artists are barbers or else some of the barbers are artists.
None of the barbers is an artist.
Is it possible for both statements to be true at the same time?

The second assertion, *none of the barbers is an artist*, is inconsistent with the mental models above, and so the theory predicts that individuals should respond that the two assertions are inconsistent. In this case, they will be correct, because the second assertion is also inconsistent with the fully explicit model above.

An example of an illusory problem based on a biconditional assertion is:

2. Illusion (biconditional)

All of the artists are barbers if and only if all of the barbers are artists.
None of the artists is a barber.
Is it possible for both statements to be true at the same time?

Biconditional assertions are true whenever both of its clauses are true or else when they are both false. In the case of a biconditional, the principle of truth predicts that a mental model of a biconditional will represent the possibility in which both clauses are true, but not the possibility in which both clauses are false. Hence, the biconditional assertion yields the following mental model in which the two sets of individuals are co-extensive:

[artist]	[barber]
[artist]	[barber]

According to the principle of truth, individuals should respond that the second assertion, *None of the artists is a barber*, is inconsistent with the first assertion, because the second assertion does not hold in the mental model above of the biconditional assertion. Mental models fail to represent the possibility in which both clauses of the biconditional are false, i.e., at least some of the artists are not barbers and at least some of the barbers are not artists. The fully explicit models would include both the model above and represent such a possibility, e.g.:

¬artist	[barber]
[artist]	¬barber

This model *is* consistent with the second assertion in the problem, and so the correct response is that the two assertions are consistent.

The same compound assertion can also yield a control problem:

2'. Control (biconditional)

All of the artists are barbers if and only if all of the barbers are artists.
Some of the artists are barbers.
Is it possible for both statements to be true at the same time?

The second assertion is consistent with the mental model, which is a correct possibility, and so individuals should respond correctly that the two assertions are consistent.

Method

Participants and procedure. 28 participants were recruited through an online platform hosted by Amazon.com. None of the participants had received any training in logic. Participants were told to take as much time as they needed to answer the questions and were asked to answer as accurately as possible.

Design and materials. Participants acted as their own controls and evaluated 18 sets of assertions (see Appendix), and each set contained one compound quantified assertion (e.g., *All the artists are beekeepers if and only if some of the beekeepers are not artists*) and one simple assertion. There were four sorts of problem: illusions of consistency (C/I), where 'C' denotes the predicted response of consistent, and 'I' denotes the correct response of inconsistent, their controls (I/I), illusions of inconsistency (I/C), and their controls (C/C). 12 of the problems were based on disjunctions, and 6 of them were based on biconditionals, for which it is impossible to have illusions of inconsistency. For each set of assertions, participants pressed one of two buttons on the screen (labeled 'yes' and 'no') to respond to the question 'Is it possible for both statements to be true at the same time?' The contents of the assertions concerned occupations (e.g., artists, beekeepers, and chemists). Each participant received the problems in a different random order. The corresponding mental models and fully explicit models are given in the Appendix.

Results

Table 1 provides the overall percentages of correct responses for the six sorts of problem. The data strongly support the predictions of the model theory.

Table 1: The percentages of correct responses for illusory and control problems in the different conditions

	Illusions	Controls
Disjunctions		
Consistent problems	36%	85%
Inconsistent problems	7%	75%
Biconditionals		
Consistent problems	43%	80%

Overall, the illusions (29% correct) were reliably harder than the control problems (80% correct; Wilcoxon test, $z = 4.49$, $p < 0.0001$), and 24 out of the 28 participants did worse on illusions than controls (Binomial test, $p < .00001$). There was no reliable difference in performance between disjunctive and biconditional consistent problems (Wilcoxon test, $z = 0.12$, $p > 0.9$). And participants succumbed to illusory inferences in both disjunctive and biconditional assertions. As the Table shows, however, a reliable interaction occurred: the illusions of consistency were more compelling than the illusions of inconsistency (Wilcoxon test, $z = 3.24$, $p < 0.002$). The control problems demonstrated that participants interpreted the sentential connectives correctly. The illusions of inconsistency likewise rule out the possibility that individuals had alternative interpretations of the connectives.

One might be tempted to argue that the results could be explained if individuals interpreted the exclusive disjunctions as inclusive ones. Yet this cannot be the case, because inclusive disjunctions merely add possibilities, and so they do not change the consistency of the assertions in the problems. We conclude that illusions of consistency in quantified assertions are a robust phenomenon.

General Discussion

Our results show that robust illusions of consistency, and of inconsistency, occur with compound assertions that consist of quantified clauses. Participants were more likely to succumb to illusions of consistency – they judged that assertions were consistent when in fact they were inconsistent. One contributory factor may have been that individuals need to show that no model exists in which the assertions hold in order to establish inconsistency. In contrast, to establish consistency, they only need to construct one model in which the assertions hold. That is, inconsistency calls for a more exhaustive search than consistency.

In general, illusions serve as a litmus test for mental models, because no other current alternative theory can predict or explain the results. The results of the present study support the principle of truth., They also show that judgments of inconsistency are more difficult than judgments of consistency (Johnson-Laird, Girotto, & Legrenzi, 2004). Theories based on formal rules of inference (Braine & O'Brien, 1998; Rips, 1994) cannot account for the illusions, because these theories rely on valid rules of inference. If such theories incorporated invalid rules to explain illusions, they would predict many inferences that individuals never make. Invalid inference rules are a recipe for irrationality, and could render theories of deduction unstable. Moreover, performance can be enhanced when reasoners are given remedial instructions (Khemlani & Johnson-Laird, 2009; Yang & Johnson-Laird, 2000).

Theories based on the probability calculus cannot readily account for performance in the task of evaluating consistency either. Chater and Oaksford (1999, 2007) assign probabilistic meanings to quantified clauses. For instance, *All the A are B* is interpreted as meaning $p(B|A) = 1$, and *Some of the A are not B* means that $p(B|A) < 1$. But, how does one assess consistency? If it is simply a matter of consistent conditional probabilities, then a problem based on these assertions:

All of the A are B or else all the B are A.
Some of the A are not B.

should be judged as consistent, because both assertions can have a probability > 0. But, the model theory predicts that these assertions should be evaluated (erroneously) as inconsistent, and indeed 61% of our participants corroborated this prediction. At the very least, the probabilistic theory needs to add some additional machinery to cope with inconsistency. A conditional probability of the form:

$p(B \& C \& D \mid A)$

has the value of zero in case the conjunction of B, C, and D, is inconsistent, even if the individual conditional probabilities $p(B \mid A)$, $p(C \mid A)$, $p(D \mid A)$ all have non-zero values.

In sum, reasoners in our study made systematic errors in reasoning about the consistency of disjunctions and biconditionals of singly quantified assertions. They tended to err on problems that called for them to take into account possibilities that rendered the assertions false, and thus corroborated the model theory's principle of truth.

Acknowledgments

This research was supported by a National Science Foundation Graduate Research Fellowship to the second author, and by National Science Foundation Grant No. SES 0844851 to the third author to study deductive and probabilistic reasoning. We thank Olivia Kang, Cathy Haught, Matt Johnson, Sam Glucksberg, Adele Goldberg, and Laura Suttle for their helpful ideas and assistance.

References

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

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38.*

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

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

Johson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Psychology Press.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111,* 640-661.

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

Khemlani, S., & Johnson-Laird, P.N. (2009). Disjunctive illusory inferences and how to eliminate them. Memory & Cognition, 37, 615-623.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*. Oxford University Press.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world*. Hove, UK: Psychology Press.

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

Rips, L.J. (1997). Goals for a Theory of Deduction: Reply to Johnson-Laird. *Minds and Machines*, 7, 409–424.

Yang, Y., Johnson-Laird, P.N. (2000). How to eliminate illusions in quantified reasoning. *Memory & Cognition*, 28, 1050-1059.

Appendix

The problems in the experiment in an abbreviated form, their mental models and their fully explicit models

Forms of Premises and Questions	Mental Models	Fully Explicit Models	% Correct
1. All of the A are B or else some of the A are B	[A] B A B [A] B A B	A B A ¬B ¬A B	
Some of the A are not B		Illusion of consistency	11
None of the A is a B		Control “no” response	82
2. All of the A are B or else some of the B are A.	[A] B B A [A] B B A	A B A ¬B ¬A B	
All of the B are A		Illusion of consistency	0
None of the A is a B		Control “no” response	71
3. All of the A are B or else all of the B are A.	[A] B [B] A [A] B [B] A	[A] B [B] A [A] B [B] A ¬A B ¬B A	
Some of the A are not B		Illusion of inconsistency	39
Some of the A are B		Control “yes” response	68
4. Some A are not B or else some B are not A.	A B (A) [B] (B) [A] [B] [A]	[A] B [B] A [A] B [B] A ¬A B ¬B A	
All of the B are A		Illusion of inconsistency	32
Some of the A are B		Control “yes” response	96
5. None of the A is a B or else some of the A are not B	[A] A [B] (A) [B] [B]	A [B] A ¬B	
None of the A is a B		Illusion of consistency	11
All of the A are B		Control “no” response	71
All of the B are A		Illusion of inconsistency	36
Some of the B are A		Control “yes” response	89
6. None of the A is a B if and only if some of the B are not A	[A] [B]	A [B] A	
All B are A		Illusion of inconsistency	39
Some A are not B.		Control “yes” response	71
7. All of the A are B if and only if all of the B are A	[A] [B] [A] [B]	[A] [B]	
None of the A is a B		Illusion of inconsistency	50
Some of the A are B		Control “yes” response	71
8. Some A are not B if and only if some B are not A.	A ¬B ¬A B A B	[A] [B]	
All of the A are B		Illusion of inconsistency	39
Some of the A are B		Control “yes” response	96