Cognitive Science 28 (2004) 621–631

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Reinstatement, floating conclusions, and the credulity of Mental Model reasoning

Jean-François Bonnefon

Laboratoire Dynamiques Sociocognitives et Vie Politique, Université Toulouse-2, Maison de la Recherche, 5 Allées Antonio Machado, 31058 Toulouse Cedex 5, France

Received 21 May 2003; received in revised form 11 February 2004; accepted 2 March 2004

Available online 18 May 2004

Abstract

Johnson-Laird and coworkers' Mental Model theory of propositional reasoning is shown to be somewhere in between what logicians have defined as "credulous" and "skeptical" with respect to the conclusions it draws on default reasoning problems. It is then argued that in situations where skeptical reasoning has been shown to lead to problematic conclusions due to not being skeptical enough, the bolder Mental Model theory will likewise make counterintuitive predictions. This claim is supported by the consideration of two of those situations, namely problems involving reinstatement and floating conclusions. It is discussed how the recent "principle of pragmatic modulation" could be a first step in order to overcome the mild credulity of Mental Model reasoning.

Keywords: Default reasoning; Mental models; Credulity; Skepticism; Reinstatement; Floating conclusions

1. Introduction

Mental Model theory (MMT) has been criticized in the past for its inability to account for defeasible reasoning—that is, reasoning to retractable conclusions. Most notably, Oaksford and Chater (1991, 1993) have argued that MMT, as a semantic method of proof, was doomed to give the wrong answers to defeasible reasoning problems—even if it was made computationally tractable enough to give any answer at all. Nevertheless, as pointed out by Garnham (1993, p. 62), "it has always been assumed that the mental models theory can be extended

E-mail address: bonnefon@univ-tlse2.fr (J.-F. Bonnefon).

doi:10.1016/j.cogsci.2004.03.002

to cover everyday, defeasible, reasoning (e.g., Johnson-Laird, 1983, Chap. 6; Johnson-Laird & Byrne, 1991, Chap. 9).” And indeed, Johnson-Laird and Byrne (2002, p. 659) have argued that MMT may provide a more psychological approach to default reasoning than the formal systems devised by logicians. In particular, Johnson-Laird and Byrne (2002) have specified how a default rule is to be represented in MMT, a topic to which I will return shortly.

The purpose of this article is to characterize MMT from the point of view of default logics, and to consider some implications of this characterization. Specifically, I will show that the inference from mental models can be classified as somewhere in between *credulous* and *skeptical*, and is thus likely to draw the wrong conclusions in situations where skeptical inference itself has been shown by Horty (2001, 2002) not to be skeptical enough. I will then consider two examples of such a situation, namely problems involving *reinstatement* and *floating conclusions*. This article will thus complement Oaksford and Chater's (1993) argument by showing how (and why) MMT derives counter-intuitive conclusions on defeasible reasoning problems featuring several extensions. My account also concurs with Oaksford and Chater's suggestion that MMT needs a more elaborated account of knowledge storage and retrieval in order to overcome its difficulties with default reasoning, an account towards which the recent principle of pragmatic modulation (Johnson-Laird & Byrne, 2002) is but a first step.

2. Mental models: a reminder

The Mental Model theory of propositional reasoning (Johnson-Laird, Byrne, & Schaeken, 1992) is subject to an ongoing scientific debate (e.g., Bonatti, 1994; Johnson-Laird, Byrne, & Schaeken, 1994; O'Brien, Braine, & Yang, 1994). From these contributions and others (such as Bell and Johnson-Laird's (1998) extension of MMT to modal reasoning, and Johnson-Laird and Byrne's (2002) Mental Model theory of conditionals) some basic tenets of the theory can be agreed upon (I will concentrate on the aspects that are relevant to the point I will make—leaving aside, e.g., the issue of fleshing out implicit models).

2.1. Step one: turning premises into models

Reasoners use premises and their general knowledge to build mental models, that is, possibilities under consideration. Models only represent what is considered to be true. A conjunction “A and B” is represented as a single model:

A	B

whereas a disjunction “A or B” is represented as three models:

A	
	B
A	B

Models may contain negated propositions; the conjunction "A and not-B" will be represented as the following, single model:

A ¬B

A conditional "If A, then B" is represented by the following pair of models, where the three-dot model denotes the presence of implicit models that contain not-A:

A B

. . .

Johnson-Laird and Byrne's (2002, pp. 659–660) discussion of default conditionals such as "If a match is struck properly, then it lights" (providing it is not wet, there is oxygen in the room, etc.) makes it clear that such assertions are initially represented by the pair of models above, that is, with respect to this example:

match–struck *match–lights*

. . .

Following Johnson-Laird and Byrne (2002), all default conditionals "If A, then B" will be represented in the remaining of this article as the pair of models:

A B

. . .

2.2. Step two: combining the models

Initial models built to represent a situation are combined to yield a final set of models, according to the following rules:

1. If two models are implicit (three-dot models), their combination results in an implicit model.
2. If one is explicit and the other implicit, no new model is formed out of their combination.
3. If two models are inconsistent (one contains a proposition and the other its negation), no new model is formed out of their combination.
4. Otherwise, the two models are conjoined together, eliminating any redundancies.

Let us take the example of two premises, "If A then B" and "A." The conditional premise is represented as the pair of models:

A B

. . .

The categorical premise is represented as a single model:

A

The model of the categorical premise is then combined with the first and second models of the conditional. No model is formed out of the combination of the model A and the implicit model "... " (from rule 2), and the combination of model A and model A B yields the model A B (from rule 4). Thus, the final model of premises "If A then B" and "A" is:

A B

2.3. Step three: deciding on a conclusion

When a final set of models has been reached through the combination of initial models, a proposition is judged: (a) *Necessarily false* (and not endorsed) when all models contain its negation and, therefore, no model contains its assertion; (b) *Possible* (but not endorsed) when at least one model contains its assertion and some models contain its negation; (c) *Endorsed* when at least one model contains its assertion and no model contains its negation; and (d) *Necessarily true* (and endorsed) when all models contain its assertion and, therefore, no model contains its negation.

As we have just seen, premises "If A then B" and "A" yield the final model A B . From this final model, one can reach the conclusion that both "A" and "B" are necessarily true.

3. Extensions, skepticism, credulity, and mental models

Default logicians have introduced a distinction between *credulous* and *skeptical* approaches to defeasible reasoning (see Touretzky, Horty, & Thomason, 1987, for the first appearance of this terminology). This distinction relates to situations wherein a given set of premises (including one or several default conditionals) has several *extensions*, i.e., several internally consistent sets of conclusions, each set nevertheless being inconsistent with the others. I will begin by defining the credulous/skeptical opposition; then I will consider whether MMT can be classified as a skeptical or as a credulous approach to default reasoning.

3.1. Skeptical and credulous conclusions: is Nixon a pacifist?

Typically, when one is reasoning from incomplete information and exception-flawed rules (which is often the case with everyday reasoning), one is led to consider not one but several sets of possible conclusions, the *extensions* introduced above. The well-known "Nixon Diamond" is an illustration of such a situation:

(1) If one is a Quaker, then one is a Pacifist,
(2) If one is a Republican, then one is not a Pacifist,
(3) Nixon is both a Quaker and a Republican.

This problem has two extensions, one where Nixon is a Quaker, a Republican, and a Pacifist, and another where Nixon is a Quaker, a Republican, but is not a Pacifist. The problem is then to decide which conclusions can be retained from those different extensions. Logicians have defined two ways to achieve this, the credulous way and the skeptical way.

The credulous way consists in randomly choosing an extension and accepting all the conclusions in this extension. Thus, a credulous reasoner confronted with Nixon Diamond will randomly choose between the conclusion that Nixon is a pacifist and the conclusion that he is not. Skeptical reasoning, on the contrary, only allows derivation of the conclusions which are present in all the extensions. Hence, skeptical reasoning on the Nixon Diamond only allows for the reiteration that Nixon is both a Quaker and a Republican, but not for any conclusion about his pacifism.

3.2. Is MMT credulous or skeptical?

As a preliminary, it will be useful to clarify the relations between models and extensions. A given set of premises involving default assertions will, on the one hand, be represented by a set of models; on the other hand, it will have a number of extensions. Let us say that an extension matches a model when both contain the exact same propositions. *Any extension matches either one model or the conjunction of several models, and any model matches a subset of one or several extensions.*

Let us consider as an example the three default rules "If A then B," "If A then C," and "If D then not-B," and the two propositions "A" and "D." This set of premises has two extensions, E1 = {A, B, C, D} and E2 = {A, not-B, C, D}. It is, on the other hand, represented by the following, final set of three models, M1, M2, and M3:

(M1)	A	B	D
(M2)	A	C	D
(M3)	A	D	¬B

The first extension, E1, matches the conjunction of M1 and M2, while E2 matches the conjunction of M2 and M3. M1 matches a subset of E1, M2 matches a subset of both E1 and E2, and M3 matches a subset of E2.

Since MMT requires for a conclusion to be drawn that this conclusion is not denied in any model, it cannot be said to be entirely credulous. Credulous reasoning makes it possible to accept a conclusion even if it is denied in a given extension—now if this conclusion is denied in a given extension, then it is denied in a given model, and thus cannot be a conclusion of MMT. For example, the conclusion that Nixon is a pacifist (or, indeed, that he is not) is not endorsed by MMT. Thus, MMT does not endorse all the conclusions a credulous approach could.

On the other hand, all skeptical conclusions are endorsed by MMT. A skeptical conclusion is present in all the extensions of a set of premises, and thus not denied in any. As a consequence, it is present in at least one model, and not denied in any—it is therefore a conclusion of MMT. Thus, MMT endorses all the conclusions a skeptical approach does.

This last observation is actually all we need to move on to the rest of this article. Yet, I wish to point out that *MMT endorses some conclusions that are not skeptical conclusions.* Such a situation arises when a conclusion, while not present in every extension, is still present in one model and not denied in any, as in the following example.

Let us consider three default rules "If A then B," "If B then C," "If D then not-B," and two propositions "A" and "D." This problem has two extensions, E1 = {A, B, C, D} and E2 = {A, not-B, D}. Note that "C" is not present in E2, and is therefore not a skeptical conclusion.

Now consider the mental models of those same premises. According to Santamarìa, Garcìa-Madruga, and Johnson-Laird (1998, p. 104), MMT predicts that the first two premises, “If A then B” and “If B then C” will initially be represented as the following pair of models:

A B C

. . .

Premise “If D then not-B” is represented as the following pair of models:

D ¬B

. . .

After the models of categorical premises “A” and “D” are combined with those four models, the final set of models is:

A B C D

A D ¬B

Since C is present in the first model and not denied in any, it is a conclusion that MMT will draw, while it is not a skeptical conclusion. *Thus, while MMT cannot be said to be strictly credulous, it is still less skeptical than skeptical reasoning proper*. Now even skeptical reasoning has been shown by Horty (2001, 2002) to be too bold an approach on some default reasoning problems. Accordingly, it is predictable that the mildly credulous MMT will likewise be too bold in those situations where skeptical reasoning proper is not skeptical enough. I will now consider in turn two such situations.

4. Floating conclusions: Mark, Lisa, and Emma—how many times have they met?

Floating conclusions have been defined within the framework of *argument systems* rather than default logic. Argument systems operate on argument extensions rather than statement extensions; accordingly, within argument systems, skeptical or credulous reasoning is defined in reference to argument extensions rather than statement extensions. However, since going any further into argument systems in this article would lead to unnecessary complications, my present purpose will only be to show that MMT gives counter-intuitive answers to problems involving floating conclusions or reinstatement, not to show in detail how argument systems operate on such problems. Besides, the examples I will use below are simple enough (in terms of argument complexity) to be handled in a quasi similar fashion by default logic and argument systems. Thus, I will contend myself with considering for each problem (a) its statement extensions, (b) its mental models, (c) its skeptical conclusions within default logic, and (d) its MMT conclusions.

Consider the following problem (all conditionals are default assertions):

(4) If A then C and D,

(5) If B then not-C and D,

(6) A and B.

Within the framework of argument systems, “D” is called a floating conclusion because it is supported in all argument extensions of the problems, but by different arguments. This conclusion is also present in all statement extensions of the problem, which are E1 = {A, B, C, D} and E2 = {A, B, not-C, D}. “D” is thus a skeptical conclusion of premises (4)–(6). Let us now consider the mental models of premises (4)–(6).

Premise (4) is initially represented by the following models, where the three-dot model stands for implicit models containing *not-A*:

A C D

. . .

Premise (5) is initially represented by the following models, where the three-dot model stands for implicit models containing *not-B*:

B ¬C D

. . .

When the model A B of the conjunction in premise (6) is incorporated into the models above, premises (4)–(6) yield the following final models:

A B C D

A B ¬C D

The conclusion “D” is present in all final models of premises (4)–(6). Therefore, not only the conclusion “D” is predicted by MMT, but it is necessarily true according to the rules for deciding on a conclusion. Skeptical reasoning and MMT thus concur in accepting “D” as a conclusion, a conclusion which would even be necessarily true according to MMT. Yet, this conclusion can sometimes be hardly acceptable, as in the following story.

Imagine that Mark and Lisa (a couple) are wondering about how many times they have met with a certain Emma. Mark and Lisa are usually reliable: as a default rule, what they say is usually the case. Mark claims they have met with Emma only once, at Emma’s. Lisa claims they have met with Emma only once, at Mark and Lisa’s. From these two claims, is it certain that Mark and Lisa had only one meeting with Emma? Such a conclusion seems hardly acceptable. The fact that Mark and Lisa contradict each other on the place this single meeting took place casts doubt on the accuracy of their memories. Although they both claim they have met only once with Emma, they could both be wrong in this respect. However confident we may feel that they had only one meeting with Emma after having listened to Mark, our confidence is likely to decrease after having listened to Lisa. In any case, judging that it is necessarily true that they have only met once with Emma seems too bold a conclusion.

However, MMT does predict it is necessarily true that they have only met once, as it will be clear if “A,” “B,” “C,” and “D” in premises (4)–(6) by are replaced by “Mark claims that they have met only once, at Emma’s,” “Lisa claims that they have met only once, at Mark and Lisa’s,” “They have met at Emma’s,” and “They have met only once.” (The proposition “not-C,” i.e., the negation of meeting at Emma’s, is here restricted to meeting at Mark and

Lisa's. This is of no consequence to the point I am making.) The final models of premises (4)–(6) then become:

Mark_claims *Lisa_claims* *Met_only_once* *Emma's*
Mark_claims *Lisa_claims* *Met_only_once* *Mark_and_Lisa's*

All final models feature *Met_only_once*. Hence, it follows with certainty from MMT that Mark and Lisa met only once with Emma. This is a first example of a situation wherein the lack of skepticism of MMT leads to a counterintuitive conclusion. I am now going to turn to another example, inspired by Horty's (2001) examination of the phenomenon of reinstatement.

5. Reinstatement: is Beth a millionaire?

The notion of reinstatement has also been developed within the framework of argument systems. To endorse reinstatement is to count an argument as acceptable even if it is defeated, as long as all the arguments defeating it are themselves strictly defeated. Once again, it will not be necessary to go further into this definition, as the example I am going to use will only be considered from the point of view of default logic on the one hand and MMT on the other hand. My purpose in using this example will only be to show how MMT derives a counter-intuitive conclusion on a problem known to pose a similar threat to skeptical reasoning, may this skeptical reasoning operate on statement or argument extensions.

Consider the following problem (all conditionals are default assertions):

(7) If A then B,
(8) If A and C then D,
(9) A and C and not-D.

Premises (7)–(9) have only one extension, E = {A, B, C, not-D}. All propositions in this extension (and, in particular, "B") are skeptical conclusions of premises (7)–(9). From the perspective of MMT, the initial models of premise (7) are the following, where the three-dot model stands for implicit models containing *not-A*:

A B
...

The initial models of premise (8) are the following, where the three-dot model stands for implicit models containing either *not-A* or *not-C*:

A C D
...

When the model A C ¬D of premise (9) is incorporated into the initial models of premises (7) and (8), only one model remains. The three-dot models are eliminated when combined with A C ¬D (according to rule 2 for combining models), and the combination of model A C ¬D and model A C D from premise (8) yields no model (according to rule 3). Finally, model A C ¬D and model A B from premise (7) are conjoined to yield the final model of premises (7)–(9):

A B C ¬D

Thus, premises (7)–(9) are represented by a single model, which matches the single extension of premises (7)–(9) in default logic. According to the rules for deciding on a conclusion, all propositions in this model should be necessarily true, including the conclusion "B." We will now consider an example (adapted from Horty, 2001) showing how this conclusion "B" may sometimes be counter-intuitive.

Let us consider that, as a default rule, employees of corporation Alpha are millionaires. We also know, as a default rule, that *junior* employees of Alpha are poor. Now we learn of Beth, who is a junior employee of Alpha, and who is not poor. Is Beth a millionaire? The sensible answer seems to be that she may, or may not be. The conclusion that Beth is a millionaire for sure is clearly too bold. Yet, such is the prediction of MMT, as shown by replacing "A," "B," "C," and "D" in premises (7)–(9) by "being an employee of Alpha," "being a millionaire," "being a junior employee," and "being poor." The final model of premises (7)–(9) then becomes:

employee_Alpha millionaire junior ¬poor

Since this single model features *millionaire*, MMT predicts that the conclusion "Beth is a millionaire" is necessarily true. This is a second example of a situation wherein MMT, as expected, suffers from some lack of skepticism: When skeptical reasoning itself is not skeptical enough, the slightly bolder MMT is prone to make counter-intuitive predictions.

6. Conclusion: a way out?

I have made the point that, being somewhere in between credulous and skeptical, MMT is doomed to lead to counterintuitive conclusions on problems where skeptical reasoning itself is not skeptical enough. This point has been illustrated by two examples drawing onto Horty's (2001, 2002) analyses of reinstatement and floating conclusions. In this final section, I will first consider the possibility that MMT may actually make the correct predictions on the examples I have considered, that is, that however counter-intuitive, they may still be the conclusions chosen by human reasoners. Leaving opened this empirical question, I will then consider possible ways for MMT to fix its mild credulity, if indeed this mild credulity is seen as a problem.

I have simply appealed to the intuition of the reader when judging that Mark and Lisa meeting Emma only once and Beth being a millionaire were inappropriate conclusions. Yet, a defender of MMT may argue that were human reasoners to actually endorse those conclusions, a point would be made in favor of the descriptive value of MMT, if not in favor of its normative value. This is an empirical question, which lies outside the scope of this brief article: My point was to show that MMT could be classified as somewhere in between credulous and skeptical, and that, as a consequence, it would endorse even the boldest conclusions of skeptical reasoning, which are sometimes bold enough to verge on the counter-intuitive side.

Now if indeed the mild credulity of MMT is seen as a problem to be fixed, one may consider two routes thereto: (a) to change the way MMT derives conclusions in order to make it more skeptical; or (b) to refine the pragmatic dimension of the theory so that it is able do differentiate between situations where floating conclusions or reinstated arguments are acceptable and situations where they are not.

We have seen that MMT always draws all skeptical conclusions, plus, sometimes, a number of other conclusions. A first step towards skepticism would be to change the rules for representation, combination, or decision so that MMT draws all but only skeptical conclusions. This would not, however, solve the problems I have highlighted in this article, since Beth being a millionaire and Mark and Lisa meeting only once with Emma are themselves skeptical conclusions. The next step would then be to turn MMT from skeptical to "over skeptical," by further restraining the set of MMT conclusions to a subset of skeptical conclusions. It is however unclear what such a term would mean for the theory, let alone how such a transformation could be achieved.

The solution of choice would therefore be to refine the pragmatic component of the theory. Reinstatement and floating conclusions are not always problematic: Only on a minority of problems do these patterns of reasoning lead to untoward conclusions. Therefore, it may be possible to understand the specificity of these problems, a specificity that probably lays in the presence of hidden, implicit information (see indeed Prakken, 2002, for such a suggestion).

As for now, the most advanced pragmatic feature of MMT is its "principle of pragmatic modulation":

> The principle of pragmatic modulation: the context of a conditional depends on general knowledge in long-term memory and knowledge of the specific circumstances of its utterance. This context is normally represented in explicit models. These models can modulate the core interpretation of a conditional, taking precedence over contradictory models. They can add information to models, prevent the construction of otherwise feasible models, and aid the process of constructing fully explicit models. (Johnson-Laird & Byrne, 2002, p. 659)

The very general nature of this principle makes it difficult to decide how exactly it should apply to a given problem: Trying to apply the principle of pragmatic modulation seems to amount to, most of the time, relying on ad hoc considerations. But ad hoc considerations on the way reasoners interpret the Beth story or the Emma story would not provide a satisfactory way out for MMT—in particular, we should not simply state that reasoners accept a conclusion only when it is consistent with background knowledge of theirs, for this would shift the burden of explanation from the theory itself to some unspecified, extraneous knowledge processing mechanism. As pointed out by Oaksford and Chater (1993) in their discussion of semantic methods of proof (including MMT) and defeasible reasoning, appeals to content, plausibility or relevance would only assume here what is to be explained.

In its present state, the principle of pragmatic modulation cannot help to address the issues I have raised here. Nevertheless, if a solution to those issues is to be found, it will most certainly emanate from a systematic development of the pragmatic, interpretative component of MMT. It is likely that we will consider the principle of pragmatic modulation to have been the first step in this development.

Acknowledgments

I wish to thank the following individuals for their helpful comments on drafts of this article: Nick Chater, Rui Da Silva Neves, Denis Hilton, John Horty, Phil Johnson-Laird, Mike Oaksford, and Eric Raufaste.

References

Bell, V., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science, 22*, 25–51.
Bonatti, L. (1994). Propositional reasoning by model? *Psychological Review, 101*, 725–733.
Garnham, A. (1993). Is logicist cognitive science possible? *Mind* & *Language*, *8*(1), 49–71.
Horty, J. (2001). Argument construction and reinstatement in logics for defeasible reasoning. *Artificial Intelligence and Law, 9*, 1–28.
Horty, J. (2002). Skepticism and floating conclusions. *Artificial Intelligence*, *135*, 55–72.
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove: Lawrence Erlbaum.
Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning pragmatics and inference. *Psychological Review*, *109*, 646–678.
Johnson-Laird, P. N., Byrne, R. M. J., & Shaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, *99*, 418–439.
Johnson-Laird, P. N., Byrne, R. M. J., & Shaeken, W. (1994). Why models rather than rules give a better account of propositional reasoning: A reply to Bonatti and to O'Brien, Braine, and Yang. *Psychological Review*, *101*, 734–739.
Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind* & *Language*, *6*(1), 1–38.
Oaksford, M., & Chater, N. (1993). Logicism, mental models and everyday reasoning: Reply to Garnham. *Mind* & *Language*, *8*(1), 72–89.
O'Brien, D. P., Braine, M. D. S., & Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review*, *101*, 711–724.
Prakken, H. (2002). Intuitions and the modelling of defeasible reasoning: Some case studies. In S. Benferhat & P. Besnard (Eds.), *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning, NMR'2002* (pp. 91–99). Morgan Kaufmann.
Santamaría, C., García-Madruga, J. A., & Johnson-Laird, P. N. (1998). Reasoning from double conditionals: The effects of logical structure and believability. *Thinking* & *Reasoning*, *4*, 97–122.
Touretzky, D., Horty, J., & Thomason, R. (1987). A clash of intuitions: The current state of nonmonotonic multiple inheritance systems. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence, IJCAI'87* (pp. 476–482). Morgan Kaufman.