Mental metalogic and its empirical justifications:
The case of reasoning with quantifiers and predicates

Yingrui Yang [1,3] and Selmer Bringsjord [1,2,3]
yangyri@rpi.edu ● selmer@rpi.edu
[1]Department of Cognitive Science
[2]Department of Computer Science
[3]Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute, Troy, NY 12180 USA

Abstract

We report on a series of experiments designed to test a new, unifying theory in psychology of deductive reasoning: mental metalogic (MML, for short; Yang & Bringsjord, 2001), which marks the unification of mental logic (ML) and mental model (MM) theory, and stands in part on the strength of empirical investigation of strategic interactions between logical syntactic and logical semantic processes in human reasoning. MML promises to resolve the long-standing controversy between ML and MM.

I. Introduction

Few would dispute the claim that reasoning, including deduction (which is the focus of the present paper), is a fundamental human ability. But how do people reason deductively, by applying inference schemas or by constructing mental models? (In asking this question we are not begging the question against those who hold that deductive reasoning proceeds neither by the application of schemas nor by the principled use of mental models. It's just that herein we are specifically concerned with MM and ML. We return to this issue in the final section.) In the last three decades, the psychology of reasoning has produced countless papers in psychological journals (such as *Psychological Review*, *Science*, and *Behavioral & Brain Sciences*) aimed at providing an answer to this question. Alas, the answer and explanation have never arrived. We have proposed a new theory, mental metalogic (MML), which attempts to integrate current mental logic and mental models theories, and aims to study the interactions between syntactic and semantic processes in human deductive reasoning (Yang & Bringsjord, 2001, 2003).

Mental metalogic is based on *both* Braine's mental logic theory (an alternative mental logic approach is given by Rips, 1994) *and* Johnson-Laird's mental models theory. This integrated theory strikes us natural, from the standpoint of modern symbolic logic, because without either formal syntax (here taken to include proof theory, based on rules of inference) or formal semantics (which includes model theory), it would be impossible to have any metatheory — and advancing metatheory has become the essence of the field of logic.

In order to see more specifically, first, how MML accounts for empirical results that ML and MM handle, consider some of the recent psychological literature on quantified predicate reasoning from the standpoints of both ML and MM: e.g., Yang, Braine, & O'Brien, 1998; Yang & Johnson-Laird, 2000$_a$, 2000$_b$, 2001. The experimental materials in question share the same surface content (viz., attributes of beads, a group of children and its subgroups, such as boys and girls). To show that ML can accommodate this material, we conducted a large-scale empirical investigation (i.e., Yang, 1996) of Braine's mental predicate logic system (Braine, 1998). This work established a parametrical model, which predicts problem difficulties by using schema-weights generated from perceived difficulty ratings. This methodology is now well-established in the psychological literature on reasoning (e.g., see Osherson, 1975; Rips, 1994; Braine, Reiser, & Rumain, 1984; called "BRR data" for mental propositional logic). The results provided not only direct empirical evidence in support of mental logic, but also challenged mental model theory. The challenge was that most inferences tested concerned quantified *compound* statements (e.g., a quantified disjunction), but mental model theory at that time only provided initial model representations for quantified *atomic* statements.

On the mental models side, in recent years the most significant developments have pertained to illusory inference. Johnson-Laird and his collaborators have published a dozen articles on this topic, including the widely read one in *Science* (2000). Sensitive to these developments, we empirically confirmed a number of MM-based predictions regarding human illusions in quantified reasoning — predictions flowing, in large part, from the Principle of Truth (POT): viz., that people construct mental models based mostly on what is true but not what is false (Yang & Johnson-Laird, 2000$_a$). In this work, predictions about the antidotes to these illusions were specifically confirmed (2000$_b$). These results showed not only the explanatory power of mental model theory, but also seriously challenged mental logic theory. The challenge was that no co-current mental inference schemas could account for the Principle of

Truth, which can be used to predict both the controls and the illusions; the principle thus seems to be fundamentally semantic.

So, at this point there is data explained by ML, data explained by MM, and phenomena to challenge both theories. From our point of view, the challenge to each side required a solution, and we propose a fairly radical one: a new theory of reasoning: mental metalogic. The initial stage for building MML was to modify the MM and ML representations in order to account for up-to-date empirical data. A new representation scheme able to bridge the separate syntactic and semantic representational systems was needed, and is now accomplished. But in this paper we focus on part of the next, and probably harder, stage, namely, the empirical investigation of the syntactic-semantic *interactions* predicted by MML.

II. Empirical Program and Results

MML theory has been developed into a principled framework to account for the interactions between logical syntactic processes and logical semantic processes in reasoning (Yang & Bringsjord, $2001_{a,b}$). Two issues are worth addressing here, one theoretical, the other empirical. The theoretical issue concerns so-called logical rationality: What *logical expectation* in human deductive reasoning does a psychological theory imply? Mental logic theory is committed to the logical expectation that people reason strategically by applying inference schemas. Mental model theory is committed to the logical expectation that people reason by constructing mental models. As a meta-theory, MML is committed to the logical expectation that human deductive reasoning may involve strategic interactions between applying inference schemas and constructing mental models. Note that MML allows purely syntactic or purely semantic processes as its special case in modeling. The empirical issue is about the *problem types*, which we suggest is a key notion in conducting empirical research on human reasoning: What problem type can a psychological theory account for? Psychology of reasoning is an empirical science. Even in large-scale empirical research, only a limited number of reasoning problems, rather than a logical system, can be examined. (The space of problems is surely infinite.) Thus, the notion of problem type is essential to empirically test MML: What problem type would demand mixed strategies between applying inference schemas and constructing mental models? This paper will identify five problem types. The first three problem types are given immediately below. Problem Type 3 was used in Experiment Sets 1 and 2. Problem Type 4 (used in Experiment 3) and Problem Type 5 (used in Experiment 4) can both be seen as variants of Problem Type 3, as will be seen.

Problem Type 1 is a set of experimental problems used to examine one mental predicate logic theory (Braine, 1998; Yang, Braine, and O'Brien,1998). Here is an example:

> All the beads are wooden or metal.
> The wooden beads are red.
> The metal beads are green.
> The square beads are not red.
> Are the square beads green?

By mental logic theory, this is a three-step problem. Each step involves applying an inference schema: by Premises 1-3 one could infer "all the beads are red or green", from which one could infer "All the square beads are red or green", then by Premise 4 one would answer "yes" to the question.

Problem Type 2 is a set of experimental problems used to examine some illusory inference predicted by the mental model theory (Yang and Johnson-Laird 2001). Here is an example:

> Only one of the following statements is true:
>   Some of the plastic beads are not red, or
>   None of the plastic beads are red.
> Is it possible that none of the red beads are plastic?

To solve this problem correctly, one has to consider two possible cases: when the first premise is true and second premise is false, and when the first premise is false and the second premise is true. Mental model theory predicates that this is an illusory problem by the semantic hypothesis known as *Principle of Truth*: People often represent only what is true, not what is false (Yang & Johnson-Laird, 2001).

Problem Type 3 can now be constructed by simply integrating Problem Types 1 and 2, an instance of which is given below:

> The premises given below are either all true or all false:
>   All the beads are wooden or metal.
>   The wooden beads are red.
>   The metal beads are green.
>   The square beads are not red.
>   Is possible that the square beads are green?

As we pointed out in the introduction, either mental logic theory or mental model theory alone would have some difficulty accounting for this new problem type. MML claims that based on the current mental logic and mental model theories, Problem Types 3 may demands a mixed strategy between applying inference schemas and constructing mental models. Some more detailed analyses will be given in the discussions following Experiments 1 and 2.

Experiment 1

Problem Type 3 can be used to manipulate three independent variables, explained as follows. The first independent variable is about the order of true/false in the heading statement. The heading statement in the above example can also be stated as: "The premises given below are either all false or all true". The second independent variable is about the set of premises. For a given problem, it can have the set of original premises, or the denials of these premises. The third independent variable is how a question is presented. It can take the form, "Is it possible that …" or "Does it necessarily follow that …". Thus, by manipulating these three independent variables, eight sub-types for a given problem are produced. For example, consider the instance of Problem Type 3 above; its "false/true, denial, and necessity" version is:

> The premises given below are either all false or all true:
> Some of the beads are neither wooden nor metal.
> Some of the wooden beads are not red
> Some of the metal beads are not green.
> Some of the square beads are red.
> Does it necessarily follow that the square beads are green?

Design A 2x2x2 between-subjects design was used to manipulate the three independent variables in eight conditions according to the 8 problem types explained above. 18 original multi-step problems similar to the example above were carefully selected from Yang, et al. (1998; Appendix 1, where the 18 problems were listed as: 17, 18, 35-38, 40, 41, 43-45, 55-57, and 59-62) in conformity to the following criteria: (1) The conclusion deductively follows from the premises given in each problem (called "true problems" in Yang, et. al., 1998). (2) For each problem, simultaneously denying each of the given premises forms a consistent set, i.e., no contradiction ensues. (3) For none of the problems does the conclusion follow from the set of denied given premises. Each of the 18 original problems was edited into 8 problem types. Doing so yields eight sets of 18 experimental problems with the same type.

Materials. Each of these sets was printed in a booklet with one problem on each page, with enough space for the participants to make notes. The participants were given a single booklet of problems, so they encountered only one problem type in the experiment.

Procedure. The participants were tested individually in a quiet room. They were given written instructions, which included one practice problem. They were encouraged to write or to draw on the problem page whatever they had in mind during the course of solving a problem. Their task was to choose among the given responses (i.e., Yes, No, or Can't tell).

Participants. Eight samples of 20 RPI undergraduates were paid to participate in the experiment. They had no formal training in logic.

Results. The mean accuracy for each sub-problem type is given in Table 1 (follows Experiment 2). For each two problem sets with the same types of premises and question, order of true/false in the heading statements had no effect. Thus, we combined Data Sets 1&2, 3&4, 5&6, and 7&8 as listed in the table above into four data sets. The results are clear-cut. For the problems using original premises, the problem type of possibility was evaluated significantly more accurately than the problem type of necessity (Mann-Whitney $Uz = 5.17$, $p < .001$). For the problems using the denials of the original premises, the problem type of necessity was evaluated significantly more accurately than the problem type of possibility (Mann-Whitney $Uz = 5.14$, $p < .001$). In addition, there was a reliable interaction. The difference between problem types of necessity and possibility for the problems using original premises was greater than for the problems using the denials of the original premises (Mann-Whitney $U = 44$, $p < 0.01$).

Experiment 2

This is a cross-variant validation test relative to Experiment 1 by using a within-subject design. All the 18 experimental problems used in Experiment 1 involved only monadic predicates, which denotes properties such as "All the beads are red". Each problem can have a variant involving dyadic predicates, which denote relations such as "All the children found red beads". For example, a dyadic version of the above instance (of Problem Type 3) is:

> The premises given below are either all true or all false:
> All the children found either wooden beads or metal beads
> The wooden beads are red.
> The metal beads are green.
> The boys did not find red beads.
> Is possible that the boys found green beads?

A set of 20 dyadic problems (of which 18 parallel the set of monadic problems used in Experiment 1) was constructed; two additional problems were selected from the same Appendix (Yang, Braine, & O'Brien, 1998), because a within-subject design was used in Experiment 2. Five data sets were collected following a procedure similar to that employed in Experiment 1 (all the experimental sessions were individually administered). Because Experiment 1 eliminated the possibility of any order effect between the wordings "all true or all false" and "all false or all true" in the heading statements, only two independent variables were manipulated this time: original premises vs.

their denials, and possibility questions vs. necessity questions; this resulted in four sub-problem types. In collecting each of the first four data sets, the booklets included two sub-problem types (half & half). The booklets used in collecting Data-Set 5 included all four sub-problem types (5 for each). Figure 1 shows that the results are highly correlated with the results from Experiment 1.

Table 1. *The mean accuracies for Experiments 1, 2, and 4.*

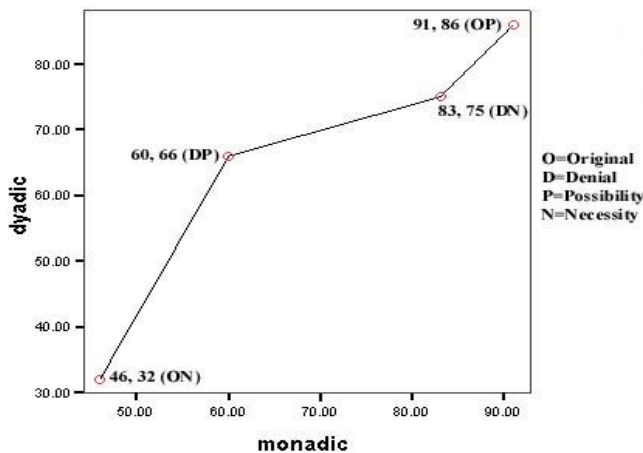|	% of Accuracy		
Problem type	Type 3 monadic	Type 3 dyadic	Type 5 dyadic
Original & Possibility	91	86	88
Original & Necessity	46	32	29
Denial & Possibility	60	66	75
Denial / Necessity	83	75	81

Figure 1.

Discussion (Exp 1 & 2)

Before we start to discuss the results, one should keep the following convention in mind. When we talk refer syntactic processes, we mean the application of some inference schemas proposed by mental logic. When we refer to semantic processes, we mean either: the application of some relevant semantic principle (such as the Principle of Truth proposed by mental model theory); the representing of a problem in mental models; or the processing of an inference in the manner proposed by mental model theory. In other, and shorter, words, the discussions will be based on the current theories of mental logic and mental models.

The results can be viewed in showing interactions of syntactic and semantic processes in a

number of ways. First, remember that all the 18 original problems are soluble by applying the inference schemas proposed by Braine's theory of mental predicate logic (1998), and reasoners can solve these problems almost errorlessly (Yang, Braine, & O'Brien, 1998). Now consider the problem type generated by using the original premises with the necessity question first. To solve problems of this type correctly, one has to consider both the case in which all the premises are true, and the one in which all the premises are false. In this situation, mental model theory, by virtue of its Principle of Truth, holds that people often fail to cope with the "all-false" case (Yang & Johnson-Laird, 2001a,b). This principle can account for the fact that 55% of participants responded "Yes" to the problems of "original-necessity" type. But notice that the illusory inference of this kind is very compelling, in the sense that in order to answer "Yes," reasoners still need to solve the problem while attending to the face value of each premise. At this "local situation," reasoners may likely apply inference schemas. However, another fairly large portion of participants (45%) responded "No" to the problems of this type, and would have needed to consider the "all-false" case. In this local situation, there are no inference schemas currently available to deal with the denials of the original premises, and reasoners may likely construct mental models.

Second, consider the problem type of using the original premises only. The results showed that participants evaluated the possibility problems much more accurately than the necessity problems. This confirmed the prediction by mental model theory that a possibility problem should be easier than its necessity counterpart; this is the case because a possibility problem requires only one model (in this case, one general model for the situation of "all the premises are true," but a necessity problem requires an additional model for the situation of "all the premises are false"). Nevertheless, after reasoners frame a single general model for the "all-true" case, by our previous point, a reasoner could still apply inference schemas in this semantically local situation.

Third, consider the problem type using the denials of the original premises only. Contrary to the previous point, the results showed that this time necessity problems were evaluated more accurately than possibility problems. It is easy to understand that for the necessity problem with the denials, there are no inference schemas that can apply, and a reasoner could construct mental models based on the immediate case in facing the denials of the original premise. However, notice that a fairly large portion of participants (60%) got the possibility problems right. In this case, some participants might guess by looking at the "all-true" case for the denials. Otherwise, a reasoner would also need to look into the "all-false" case for the denials; when this happens, which might

well be the case, the original premises would be recovered and then inference schemas can apply.

In sum, we conclude that the performance of the participants in this experiment could be heavily bounded by the semantic principles proposed by the mental model theory, but could also be explained by the syntactic capacities claimed by mental logic theory.

Experiment 3

The above discussions were largely based on the Principle of Truth, which implies that most subjects failed to consider the "all false" situations. But is this the case? Experiment 3 was designed to test this underlying hypothesis by following the same procedure as Experiment 2. The same set of experimental problems were used, except that instead of using modal questions involving possibility and necessity, this time each of the 20 problems followed an indicative question started by: Does it follow that …". We count problems of this type as *Problem Type 4*. Recall that half of the problems used original premises, from which the putative conclusion logically follows; and recall as well that half of the problems (different problems) used the denials of their original premises, from which the putative conclusion doesn't logically follow. The results were clear-cut: 75% of participants answered "Yes" to the problems with original premises, and 72% of participants answered "No" to the problems with the denials of their original premises. In the case of denials, if participants looked at the "all false" situation, their original premises could be recovered, and they would answer, "Yes", as they did in the case of originals. By the same token, participants would answer, "No", if they looked into the "all false" situation as they did in the case of denials. Thus, this result provides clear evidence in support of the proposition that the Principle of Truth was indeed used in our previous analyses.

Experiment 4

Problem Type 5 was also constructed based on Problem Type 3; the difference was that this time for each problem, a modal operator, "possibly", was inserted (once or twice) into the premises. Here is an example:

The premises given bellow are either all true or all false:
   All the children found either wooden
   beads or *possibly* metal beads
   The wooden beads are red.
   The metal beads are *possibly* green.
   The boys did not find red beads.
Is possible that the boys found green beads?

Experiment 4 was intended to enable further examination of the reliability of our discussions for Experiment 1 by using this new problem type. A set of 20 Type 5 problems were constructed parallel to the set of Type 3 dyadic problems used in Experiment 2. The same procedure as Experiment 2 was used, and the results did not show significant difference in comparison with either Experiment 1 or 2 (see Table 1). Notice that inserting modal operators into premises might change their internal logical structure, and in turn, it might cause certain framing effects. But it didn't happen. Our explanations of this are two-fold. On the one hand, it shows that the phenomena observed in Experiments 1 and 2 are very robust, and it makes us more confident about the MML analyses. On the other hand, we suspect that the result is due to the fact that Problem Types 3 and 5 share the same modal questions. (For instance, the accuracies for the sub-problem types such as the original/possibility and the denial/necessity were high, which did not leave much room to improve. An interesting experiment for further research would be using a new problem type integrating Problem types 4 and 5, which is Problem Type 6 with modal premises and indicative questions.)

III. Brief reply to potential objections

Space constraints preclude anticipating and rebutting the inevitable objections, but we will say a few words. The current version of mental metalogic theory is based in significant part on the current schism between mental logic and mental model theories; we refer to this dichotomy as the "paradigm." In general, objections to MML will come from both inside and outside this paradigm. From the inside, some mental logicians and/or mental model theorists will resist an integrated approach. They will insist that one theory (either ML or MM) is sufficient to explain all the relevant data. A detailed reply to this position includes a discussion of problem types that are solvable – whether tackled by human or machine – only through reasoning that is at once syntactic and semantic; for an example see (Bringsjord & Yang, forthcoming). In these problems, on the one hand solutions require reasoning that proceeds in step-by-step fashion using schemas, yet on the other hand the schemas reference not only linguistic patterns, but also models. We know that explicitly teaching this kind of hybrid reasoning can produce reasoners who are immune the Wason selection task and the various problems taken to show that human's reasoning skill can be largely improved through training of hybrid reasoning strategies. (Rinella, Bringsjord, Yang, 2001). What about from *outside* the mental logic/model clash? One objection from outside the paradigm might be that both mental logic and mental model theories are true (despite the fact that those inside the paradigm energetically repeat that the theories are outright competitors), and MML is just a simple combination of the two. Hopefully the empirical work reported above shows that such an objection is mistaken. Some reviewer had a more

serious charge on MML that mental logic and mental model theories are indistinguishable, referring to Stenning and Yu (1997) and other logicians. To our opinion, this is even more seriously mistaken. Briefly, three points are worth noticing here. First, Stenning and Yu (1997) only studied the case of simple syllogisms, while the problem structure used in our experiments are more complex. Second, there are certain limitations for using graphical method such as Vann Diagrams to represent complex compound statements. Third, one the one hand, even with Smullyan tree method, we don't think logicians would neglect the distinction between formal syntax and formal semantics; on the other hand, as we pointed our earlier, at the mental representation level, mental logic can be seen as a *partial* selection of formal syntax in logic and mental models as *partial* selection of formal semantics, and thus there is not current tree method to account for the function to integrate the two partial selections. Fourth, The present authors are highly respect and value Stenning and Yule proposal. We regard that the determination from choosing between their program and our MML program is an empirical question, as both programs are hypothetical about mental presentations and mechanisms; and finally, we do appreciate reviews' points and are thankful for offering us a chance to report the empirical work on MML. As to those who accept neither ML nor MM, sustained debate will be required. We plan to defend the mental logic/model paradigm against those taking a thoroughgoing probabilistic, heuristic, or verbal approach – approaches we cluster together under "eliminativism." Explicit reasoning, whether syntactic or semantic, or whether deductive, inductive, or abductive, does seem to us to be an undeniable part of human cognition, but full treatment of this profound issue goes beyond this short paper, and we refer the reader to (Yang and Bringsjord, 2003)

Acknowledgment

References

Braine, M.D.S. (1998). Steps toward a mental predicate-logic. In M.D.S. Braine & D. P. O'Brien (Eds.), *Mental Logic*. pp. 273-331. Mahwah, NJ: Lawrence Erlbaum Associates.

Bringsjord, S. & Yang, Y. (forthcoming) "The Problems that Generate the Rationality Debate are Too Easy, Given What our Economy Now Demands," *Behavioral and Brain Sciences*.

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

Johnson-Laird, P.N., Byrne, R.M.J., and Schaeken, W.S. (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.

Johnson-Laird, P,N., Legrenzi, P., Girotto, V. and Legrenzi, M.S. (2000). Illusions in reasoning about consistency. *Science*, Vol. 288, pp531-532.

O'Brien, D.P., Braine, M.D.S., and Yang, Y. (1994) Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review, 101*, 711-724.

Rinella, K., Bringsjord, S. & Yang, Y. (2001) "Efficacious Logic Instruction: People are not Irremediably Poor Deductive Reasoners," in Moore, J.D. & Stenning, K., eds., *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum Associates), pp. 851-856.

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

Stenning, K. and Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology, 34*, 109-159.

Yang Y., Braine, M.D.S., and O'Brien, D.P. (1998). Some empirical justification of one predicate-logic model. In M.D.S. Braine & D.P. (Eds.) *Mental Logic*, pp. 333-365, Mahwah, NJ: Lawrence Erlbaum Associates.

Yang, Y., & Bringsjord, S. (2001). Mental MetaLogic: A new paradigm for Psychology of reasoning. *The Proceedings of the Third International Conference of Cognitive Science*. USTC Press: Beijing, China.

Yang, Y. & Bringsjord, S. (2003). *Mental metalogic: A new, unifying theory of human and machine reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Yang, Y. and Johnson-Laird (2000). Illusions in quantified reasoning: How to make impossible seem possible, and vice versa. *Memory & Cognition. 28 (3)*, 452-465.