Cahiers de Psychologie Cognitive
Current Psychology of Cognition
1999, 18 (1), 87-97

CONTINUING COMMENTARY ON "RATIONALITY IN REASONING" (J. ST. B.T. EVANS & D. E. OVER)

The semantics of logical connectives and mental logic

David P. O'Brien[1] and Luca L. Bonatti[2]

1. City University of New York
2. University of Paris VIII at Saint-Denis and New York University

Two broad hypotheses on deductive reasoning are currently being developed. The mental-model hypothesis claims that humans construct mental models, which are something like direct analog representations of states of affairs in the world. The mental-logic hypothesis proposes that human deductive competence depends on the existence of a system of rules that acts on discrete linguistic representations probably specified in a language of thought. At first sight the distinction between the two hypotheses seems clear: they argue for alternative systems of representations. Based on these apparent alternatives, proponents of mental models since Johnson-Laird (1983) repeatedly have advanced hosts of arguments (albeit always different ones) purportedly showing that the mental-logic hypothesis must be abandoned.

Correspondence should be sent to David P. O'Brien, Box G-1126, Department of Psychology, Baruch College/CUNY, 17 Lexington Avenue, New York, NY 10010, USA (e-mail: obrbb@cunyvm.cuny.edu).

We want to say immediately that we are not enthused about this style of debate. The reason is simple: both hypotheses at the broad abstract level just described have too many degrees of freedom to be evaluated scientifically. Johnson-Laird (1983) claimed to have given "the final and decisive blow" to the doctrine of mental logic (p. 141). In order actually to have a "final and decisive blow" against either of the two hypotheses, however, one needs a specified principled philosophical argument against its core. We do not think that such arguments exist against the core of these hypotheses (or at least against the general mental-logic hypothesis), so in our opinion the only way to judge the general hypotheses is to extract real theories out of them – theories that are meant to explain some real and clearly identifiable domain of reasoning – and judge their empirical explanatory power. It is precisely for this reason that Braine, O'Brien, and their collaborators have striven to develop a real psychological theory out of the general mental-logic hypothesis. We were therefore happy to read Over and Evans (1997b) (henceforth O&E) making much the same point. They wrote, "We wish to separate the very general frameworks or paradigms for accounting for deductive competence from particular theories formulated within these frameworks. ... We do not think that one has a theory merely because of a commitment to apply mental logic or models to the study of human reasoning. We suggest it would be much more accurate to talk about contrasting frameworks or paradigms within which particular theories, like O'Brien's, can be formulated. It is possible to provide a strong empirical test of a particular mental logic or mental models account of some given phenomena, and thus we can say that some theory of O'Brien's has been confirmed or disconfirmed by some data. But neither the mental logic nor mental models frameworks as a whole can be directly confirmed or disconfirmed" (pp. 823-824).

Bonatti (1998) made exactly the same point almost verbatim.[1] We were thus much surprised to find O&E (1997b) continuing their commentary by accusing O'Brien of making the same confusion we have striven to avoid. They wrote: "O'Brien is not attending to a distinction we tried to make in our original article and to be clearer about in later replies. ... The debate between the mental logic and mental models camps reflects the failure of authors on both sides to distinguish their framework from their particular theoretical versions of it ... O'Brien's latest commentary includes several examples of this confusion" (pp. 823-824).

1 "[The principal prediction of mental logic] is that ceteris paribus the difficulty of a problem is function of the number and difficulty of the rules involved in a proof. However, such prediction has little empirical impact if the rules, their conditions of application, and their psychological costs, are not specified. The above general prediction, as such, does not afford a quick way to refute the hypothesis. Yet, surely, specific theories based on it can be refuted – this is why it is important to discuss one specific version of the

We think that O&E's judgment depended on a misreading of O'Brien's (1997b) arguments, taking O'Brien's comments about the specific mental logic theory developed by Braine and O'Brien as comments about the more general mental-logic hypothesis, and hence finding them too unconstrained. They are not unconstrained, however, because they refer to a specific theory which is sufficiently developed to be tested easily. This misunderstanding led O&E to underestimate the power of Braine's and O'Brien's theory, and as a consequence it led them to think that mental models should be preferred over mental logic, even when the very same results and arguments to which O&E referred would lean naturally towards mental logic.

We want to clear up these misconceptions. To be as clear as possible, throughout the remainder of this article we shall refer to the general mental-logic hypothesis as MLH, and to the particular mental-logic theory presented most completely in Braine and O'Brien (1998) as MLT. Although E&O (1997a, 1997b) and O&E (1997a, 1997b) raised several criticisms of both MLH and MLT, each of which deserves discussion, for reason of space we shall concentrate on what seems to be their most serious conceptual worry: that MLH and MLT do not explain the

1 (continued).
hypothesis, such as Braine's. [...] The mental model hypothesis also is only a very general framework. It holds that reasoning involves building and inspecting models, and hence it lends itself to the "principal prediction" (Johnson-Laird et al., 1992, p. 436) that the difficulty of a problem is function of the number of models required. Yet this "principal prediction" is likewise not sufficiently specific to yield univocal empirical tests. [...] many specific theories can be developed that are consistent with the hypothesis. Hence [...] its principal prediction has little empirical bearing. Just as for the mental logic case, only fully developed mental model theories can be submitted to the tribunal of experience" (Bonatti, 1998, pp. 436-437).

meanings of connectives, whereas "Johnson-Laird at least takes logical semantics seriously in his theory" (O&E, 1997a p. 270). We will argue that, on the contrary, MLH is well suited to account for the meaning of logical constants and that moreover MLT adds important elements to understand such meaning, which puts it in a unique position with respect to any other empirical theory of reasoning. Many of the other reservations that E&O have with respect to MLH are tied to this point and should evaporate once this point is clarified.

MLH and the semantics of connectives

In a nutshell, MLH holds that the meaning of a logical connectives is determined by the inferences in which that connective appears. The thesis did not originate with us, but with logicians and philosophers such as Gentzen (1964) and Dummett (1975). Other philosophers and linguists, from Block (1986) to Sperber and Wilson (1986) have in fact extended this thesis to the whole of language and have championed what is generally called an inferential (or conceptual) role semantics, according to which the meaning of any word of language is determined by the inferences in which it appears. We hope that E&O agree that at least these logicians and philosophers take logical semantics seriously, even if E&O think MLT has not. Indeed, the inferential-role semantics program is healthy and widely accepted. Furthermore, oddly enough for E&O's criticism, inferential-role semanticists often refer to the logical connectives as a paradigmatic case of how meanings can come from patterns of implications (e.g., Dummett, 1975; Peacocke, 1986).

Whether inferential role semantics is tenable for the whole of language is a question for debate (e.g., Fodor & LePore, 1991). We take a moderate stand: we do not know what the right story for the whole of language is, and we are sensitive to the doubts that meaning in general can be specified adequately in terms of conceptual roles. However, we see no argument against the possibility that inferential semantics is indeed the right semantics for the logical part of language. If E&O know of any, we will be happy to consider them, but they have neither presented nor referred to any in their previous articles, and we therefore find their general criticism totally without support at this time.

Restricted to logic, inferential-role semantics claims that the meaning of logical connectives comes from their rules of introduction and elimination. We believe that our contribution to the debate in this area is having shown that – even by restricting inferential-role semantics to the logical part of language – a natural semantics of connectives is quite difficult to construct, and most inferential-role semanticists underestimate this difficulty. It is precisely because we take semantics seriously, however, that we endorse MLT, which specifies in detail rules that govern natural connectives in natural language and hence allows one to discuss a specific hypothesis about what their meanings are.

Perhaps there is a more serious way to proceed, but certainly it is not enough to *say* that another semantics is more serious to make the case that it *is* more serious. O&E gestured towards the need for an alternative semantic theory directly constructed in terms of possible worlds. Whereas we do not deny that such a semantics is possible, we invite O&E to develop it at the level of precision that our theory already has achieved, and then we will consider whether their proposal is preferable to ours. We expect, however, that O&E, who share with us many basic tenets about what a semantics for conditionals should achieve, will quickly reevaluate both MLH and MLT as soon as they attempt to develop a model-based semantics as a psychologically plausible theory of meaning.

MLT and the semantics of conditionals

We turn now to three specific claims of O&E (see also E&O, 1997b, pp. 402-404) against MLT: (a) MLT has no semantics for *if*; (b) the proposal in Braine & O'Brien (1991) and O'Brien (1997a, 1997b) is incompatible with the very nature of deductive rules (and thus is not even coherent with MLH); and (c) MLT would make the application even of modus ponens impossible. We will show that all the three points are mistaken.

First, however, note an important point of agreement between us and E&O; like us, they hold that not any theory of conditionals will do as a psychological theory for *if*. For example, whereas from a formal perspective a model-theoretic treatment of conditionals is adequate even if it results in entailment of inferences such as the paradoxes of material implication, E&O agree with us that a psychological theory that licenses the paradoxes is not an adequate characterization of the natural meaning of *if* (see O&E, 1997a, p. 269). Johnson-Laird's models theory of con-

ditionals, however, which is essentially an abridged version of the truth-tables semantics, does license the paradoxes, as well as many other inferences that seem quite unnatural to ordinary intuitions. This makes the current state of the models theory inadequate on grounds that both we and E&O have stated, making E&O's stated preference for the models hypothesis quite mysterious.

Now let us turn to MLT. Because E&O's worry that MLH in general cannot offer a semantics for connectives is unmotivated, as we have shown above, we consider E&O's serious doubts regarding the specific implementation of MLH offered by MLT. They argued that, "Ironically, given O'Brien's vigorous claim on behalf of mental logic, this system is not strictly a mental logic at all as far as its rules for the conditional are concerned" (p. 827). Their reason for this statement is that MLT holds that in reasoning under a supposition one cannot appeal to propositions that would be false given that supposition. They make it clear that this paradoxical failure to be faithful mental logicians leaves MLT without a logical semantics for conditionals, stating that "really connected with this point is our complaint that O'Brien does not have a logical semantics for his conditionals" (p. 827).

These are quite puzzling criticisms to us, and seem to follow from a misunderstanding. The MLT proposal for a psychologically plausible semantics for conditionals is a version of the general inferential-role semantics for logic that we described above. What distinguishes MLT from other similar inferential-role semantics is the place assigned to the implementation procedures. For MLT, the introduction and elimination schemas alone are not sufficient to fix the meaning of connectives in isolation from how these rules actually are implemented in lines of reasoning. This is particularly important for *if*. A general property of the direct-reasoning routine of MLT is that derivations stop when contradictions are detected. Thus, whereas in classical logic everything follows from a contradiction, in MLT nothing follows from it. Note that this is a property of the system, and to state it one does not need to make reference to extralogical properties such as the meanings or truth of any sentences involved in the proofs; it is thus thoroughly consistent with MLH. Note also that this property, which is independently motivated, turns out to block the paradoxes of material implication, and hence should be a point in favor of what E&O agree to be an important step in the right direction for a semantics for *if*.

As is the case with most formal properties, however, a constraint that is not directly semantically motivated still has semantic consequences. In the case of *if*, the consequence is that if one wants to reason towards a conclusion within a conditional proof (corresponding to a case of introduction of a conditional), one cannot start with suppositions that are false given the premises. If one did, then one could generate a contradiction inside the subordinate derivation under the supposition and, because of the constraint, nothing would follow from it. This is a nontrivial semantic consequence of the syntactic constraint over the rules for *if*; it is independent of the schemas, but contributes to fix the meaning of *if* in that it has consequences for what patterns of implications are effective when the schemas are used. Once again, however, in no sense is the constraint in opposition to MLH, and MLT is strictly a mental logic, contra O&E's assertions.

O&E put forward an even stronger argument, however, claiming that the semantic consequence of the constraint leads to absurdity. According to O&E, it would even block unconstrained application of modus ponens and would render it dependent on the meaning of each sentence: "If this constraint is accepted, how can even Modus Ponens be a valid inference form?" (O&E, 1997b, p. 826). Although the charge is strong, O&E support it only with an example: "Consider a juror at a trial who, through a chain of reasoning, has reached the supposition that the defendant is innocent. Perhaps the juror has laid down as another premise that he is a monkey's uncle if the defendant is innocent. Does O'Brien's general constraint mean that this juror is not allowed to perform MP?" (p. 826).

We really doubt that the example is to the point, but let us assure O&E that the constraint leads to no such absurdity. If the juror has reached, as O&E wrote, the *supposition* that the defendant is innocent (notice it is a supposition rather than a conclusion) then the juror is performing a line of reasoning under the supposition *the defendant is innocent*. Now, let us continue with O&E by imagining that outside the supposition the juror also has accepted that *if the defendant is innocent I am a monkey's uncle*. Because there is no contradiction between the supposition and the conditional sentence that the juror has accepted, nothing blocks using that sentence under the supposition, to apply modus ponens *inside the suppositional reasoning*, and therefore to derive inside that subreasoning the sentence *I am a monkey's uncle*. It is this last sentence that is factually false, but not the original conditional

sentence *if the defendant is innocent I am a monkey's uncle*, which therefore can be appealed to within a conditional proof according to our constraint. O&E's apparent paradox depends on the fact they confuse the truth value of the simple sentence *I'm a monkey's uncle* with the truth value of the conditional *if the defendant is innocent then I am a monkey's uncle*, which is true by their own assumption. In short, they confuse the truth of a conditional with the truth of the consequent of a conditional, a slip we find particularly unfortunate in the context of a discussion about conditionals.

The error cleared away, however, let us follow O&E's juror a step further. In a line of reasoning by supposition, the juror has just come to conclude – under the supposition that the defendant is innocent – that he is a monkey's uncle. Under the reasonable assumption that he is not a monkey's uncle, the juror will have reached a contradiction inside his suppositional reasoning. The interesting question now concerns what will happen next. Will the juror be entitled to conclude that every possible proposition now follows under the supposition, as would occur from the semantics of standard logic, or will the reasoning stop with the feeling that somewhere something went wrong, either in the supposition or in some other proposition used to reach this contradiction? We suspect that if you think about it, you would choose the latter option. This, indeed, is the one predicted by the constraint of MLT, which assumes that nothing follows from a contradiction except that something has gone wrong. O&E's case thus is far from being problematic for MLT, but rather is a welcome supporting example of the motivations that led MLT to introduce the constraint that O&E challenge.

O&E's last attempt to argue that MLT has no semantics for conditionals was to show that in any case it will have to meet the same difficulties that O'Brien found make a possible-worlds semantics problematic. O'Brien (1997b) argued for the superiority of the MLT treatment of *if* over a model-theoretic approach of the sort proposed by Stalnaker, because such solutions require one to specify a similarity metric between possible worlds, and it seems that in most cases where we can use *if* correctly we have no clue about such a metric. O&E argued that the semantic consequence of the "nothing follows from a contradiction" constraint makes MLT "much more similar to Stalnaker's approach than O'Brien realizes" (p. 827). They argued that once a line of reasoning from a supposition P has started, one will have to figure out what further suppositions are plausible given P, and, they claimed, to solve this problem one has to give the same metric of similarity-plausibility among statements that O'Brien found problematic for the model-theoretic account. O&E are correct that deciding what the consequences of a supposition are often is not a logical process, but they obliterated a crucial point of difference between a model-theoretic account and MLT. In a model-theoretic account, being able to specify a metric of similarity between possible alternative states of affairs is a *constitutive condition* for the semantics of the conditionals, and this is precisely why a model-theoretic account is implausible. MLT has no such requirement; finding out what the suppositions within a line of reasoning are thus is not a constitutive condition for the semantics of conditionals: it is simply a matter of better strategy or of plausibility. Understanding of conditionals therefore is independent from the specification of similarity, whereas it is not for the model-theoretic semantics. Contra O&E, the possible-worlds and MLT approaches are crucially different with respect to the constitutive conditions for the semantics of *if*. O&E confused a claim about the semantics for "lion" with a claim about the best strategy to follow in case we encounter a lion. We think that the two levels have to be kept separate and we submit that obliterating this distinction generated much confusion.

Conclusion

In their recent writing, although they carefully and usefully pointed out problems for both sides, E&O repeatedly expressed their preference for the mental-model framework over MLH. They did so on the basis of various arguments that were partly empirical and partly principled. We have shown that their most serious argument against MLT and MLH is not substantial. Although we have no space to do it here, we hold that we have answers to each of their other arguments. We find the principled ones against MLH are inconclusive. We also find that their empirical ones against MLT, which we consider more serious, either can be addressed by the theory as it stands (e.g., their worry that the fact that modus tollens is not a basic rule is an arbitrary assumption) or point to interesting directions to develop parts of MLT that are still unexplored (e.g., their challenge to account for the negative conclusion bias within MLT). We thank O&E for having considered our theory and

having tried to show its shortcomings. All in all, we are happy to note that they found few, and we have answers to all of them.

We believe in rational debate and we thank E&O for the opportunity they gave us to clarify important aspects of our theory. We are confident that we can show that the reasons that led them to prefer the mental-model framework over MLH are not as strong as they seemed to believe. We are sure that E&O, whose work we deeply respect, will either come up with better arguments against our position or will change their negative judgment about it.

REFERENCES

Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 10,* 615-678.

Bonatti, L. L. (1998). What the mental logic-mental models controversy is not about. In M. D. S. Braine & D. P. O'Brien (Eds.), *Mental logic.* Mahwah, NJ: Lawrence Erlbaum Associates.

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98,* 182-203.

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

Dummet, M. (1975). What is a theory of meaning. In S. D. Guttenplan (Ed.). *Mind and language.* Oxford, UK: Oxford University Press.

Evans, J. St. B. T., & Over, D. E. (1997a). Rationality in reasoning: The problem of deductive competence. *Cahiers de Psychologie Cognitive/ Current Psychology of Cognition, 16,* 3-38.

Evans, J. St. B. T., & Over, D. E. (1997b). Reply to Barrouillet and Howson. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 399-405.

Fodor, J. A., & LePore, E. (1991). Why meaning (probably) isn't conceptual role. *Mind and Language, 6,* 328-343.

Gentzen, G. (1964). Investigations into logical deduction. *American Philosophical Quarterly, 1,* 288-306. (Originally published in 1935.)

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

O'Brien, D. P. (1997a). The criticisms of mental-logic theory miss their target. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 173-180.

O'Brien, D. P. (1997b). The criticisms of mental-logic theory still miss their target. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 813-822.

Over, D. E., & Evans, J. St. B. T. (1997a). Two cheers for deductive competence. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 255-278.

Over, D. E., & Evans, J. St. B. T. (1997b). The nature of mental logic: A reply to O'Brien. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 16,* 823-830.

Peacocke, C. (1986). *Thoughts: An essay on content.* Basil Blackwell.