Synthese (2021) 198:7249–7256
https://doi.org/10.1007/s11229-019-02519-9

Are the open-ended rules for negation categorical?

Constantin C. Brîncuș[1]

Received: 5 September 2019 / Accepted: 11 December 2019 / Published online: 18 December 2019

Abstract
Vann McGee has recently argued that Belnap's criteria constrain the formal rules of classical natural deduction to uniquely determine the semantic values of the propositional logical connectives and quantifiers if the rules are taken to be open-ended, i.e., if they are truth-preserving within any mathematically possible extension of the original language. The main assumption of his argument is that for any class of models there is a mathematically possible language in which there is a sentence true in just those models. I show that this assumption does not hold for the class of models of classical propositional logic. In particular, I show that the existence of non-normal models for negation undermines McGee's argument.

Keywords Open-endedness · Non-normal models · Categoricity · Logical constants

1 Conservativeness, uniqueness and open-endedness

McGee (2000, 2015) has recently argued that the formal rules of classical natural deduction uniquely determine the semantic values of the logical connectives and quantifiers if these rules are *open-ended*, i.e., if they are sound not only within a certain language, but they remain sound in any mathematically possible extension of that language.

The requirement of open-endedness is meant to supplement Belnap's (1962) criteria (*conservativeness* and *uniqueness*) that a rule should satisfy for the acceptability of the connective that it introduces. Conservativeness guarantees that the addition of a new connective creates a conservative extension of the initial language (i.e., it adds no new truths about the initial language) and *uniqueness* guarantees that a rule which introduces a new connective is such that it allows precisely one inferential role for that connective (i.e., if there are two syntactical connectives, $\#_1$ and $\#_2$, that obey the same formal rules and σ' is a sentence obtained from σ by replacing each occurrence of $\#_1$ with $\#_2$, then σ and σ' are interderivable). Harris (1982)

✉ Constantin C. Brîncuș
constantin.brincus@unibuc.ro

[1] Faculty of Philosophy, University of Bucharest, Bucharest, Romania

proved that the formal rules of natural deduction for classical propositional connectives and quantifiers do satisfy the uniqueness condition[1] and McGee (2000: p. 67) takes this result as showing that the rules of classical natural deduction "uniquely pin down the semantic role of the connectives and quantifiers", if they are open-ended. The semantic role of a sentence is taken by McGee (2000: p. 66) to be determined, "uniquely up to logical equivalence, by indicating the models in which the sentence is true."[2]

McGee's proposal could be seen as an attempt to offer what Carnap (1943) called a *full formalization* of classical logic, i.e., a formalization that uniquely represents all the semantic properties of the logical terms. In his 1943 book, *Formalization of Logic*, Carnap proved that the standard formalizations of classical propositional and predicate logic allow for non-normal interpretations, i.e., interpretations for which the calculi remain sound, but in which the logical constants have different meanings than the standard ones. The existence of such interpretations shows that the standard calculi do not fully formalize all the logical properties of the logical terms and, thus, fail in uniquely determining their meaning.

Does an open-ended formalization of propositional classical logic, however, eliminate the non-normal interpretations? In particular, since Carnap (1943: p. 84, T16-3) proved that if negation has a normal interpretation, then all the other propositional connectives also have a normal interpretation,[3] the problem that has to be analyzed is whether the open-ended rules for negation are categorical.[4] The analysis

[1] An extensive discussion on the uniqueness condition could be found, for example, in Humberstone (2011: pp. 578–630) and Došen and Schroeder-Heister (1985).

[2] McGee (2015) proposes an understanding of the semantic role of sentences in terms of possible worlds; namely, a sentence is taken to express a proposition and the latter is understood, following R. Stalnaker's account, as a set of possible worlds. He then formulates propositional rules for the propositional connectives, i.e., the counterparts of the sentential ones. For simplicity, we shall not consider this propositional approach here, but what we say below is applicable, *mutatis mutandis*, to it.

[3] In a non-normal interpretation, disjunction violates the fourth row from the normal truth table (NTT), i.e., it is true although both of its disjuncts are false (the first three rows are secured by the vI rule). This happens because vE rule does not fix the fourth row of the NTT. Nevertheless, if negation is normal, then disjunction is also normal, otherwise the Disjunctive Syllogism Rule (AvB, ~A ⊦ B) would become unsound (i.e., if "A" and "B" are false and negation is normal (thus, "~A" is true), then "AvB" cannot be true). However, since negation and disjunction form a functionally complete set of connectives, then all the other connectives will be normal.

[4] The notion of categoricity used in this paper differs from the standard notion of categoricity defined in modern model theory, where a theory T is categorical in a cardinal κ (or κ-categorical) if and only if it has exactly one model of cardinality κ up to isomorphism. The present notion of categoricity simply points out to the fact that the formal rules of deduction are compatible with truth-tables (the normal and the non-normal ones) that are not isomorphic. A precise definition of isomorphic truth-tables was given by Kalicki (1950: p. 175) by adapting Tarski (1938: p. 106)'s definition of isomorphic matrices. The main idea of the definition is that, in at least one row, for the same input, the normal truth-table gives a designated value, while the non-normal truth-table gives an undesignated one –or the other way around. A general definition of categoricity in this sense could be given following Scott's (1971: pp. 795–798) terminology: a formal system of logic is categorical if and only if the only valuation that is consistent with the syntactical relation of logical consequence in that system is the standard one. A valuation ν is consistent with a consequence relation $\vdash$ if and only if, whenever $\Gamma \vdash \sigma$, if $\nu(\varphi)=1$ for all $\varphi \in \Gamma$, then $\nu(\sigma)=1$. For a discussion of this notion of categoricity see Hjortland (2014: pp. 447–51) and Bonnay and Westerståhl (2016: pp. 726–27). It is worth mentioning, however, that the property of categoricity, in this sense, is relative to the format of the proof system. For instance, if we strengthen the proof-theoretic

of the open-endedness requirement becomes even more interesting because some of Harris' results had been already known by Carnap (1943: pp. 32–33, T8-9), who showed that, under ordinary conditions, if a propositional calculus contains two signs for negation ('$\sim_1$' and '$\sim_2$'), then for any closed sentence σ, $\sim_1\sigma$ and $\sim_2\sigma$ are syntactically interderivable and interchangeable (C-equivalent and C-interchangeable in Carnap's terms).

Nevertheless, the fact that the formal rules for negation (and for the other syntactical connectives) respect Belnap's uniqueness condition did not lead Carnap to the conclusion that these rules uniquely determine the meanings of the syntactical connectives that they introduce. On the contrary, in spite of the syntactical uniqueness results, Carnap discovered that negation and most of the propositional connectives allow for non-normal interpretations.[5] Moreover, there are also non-normal interpretations of the quantifiers even when the classical propositional connectives have only normal interpretations. In particular, there are sound interpretations of quantificational logic in which "(∀x)Fx" could be interpreted as "every individual is F, and b is G", where "b" is an individual constant. The possibility of these non-normal interpretations arises because, in the standard formalizations of quantificational logic, a universal sentence is not *deductively equivalent* with the class formed by the conjunction of all the instances of the operand. Nevertheless, the non-normal interpretations of propositional calculi will suffice for analysing the efficiency of the open-endedness requirement.

2 Are the open-ended rules for negation categorical?

Carnap (1943: Chapter C) proved that there are two mutually exclusive kinds of non-normal interpretations for propositional calculi (and in particular for classical negation): non-normal interpretations in which a sentence and its negation are both true (and, thus, all the sentences are true) and non-normal interpretations in which they are both false (and their disjunction is true). McGee (2000: p. 71) assumes that there is no model in which all sentences are true and, thus, excludes by stipulation the first kind of non-normal interpretations. However, he argues that if the rules are open-ended, then there is no model in which a sentence and its negation are both false (i.e., the second kind of non-normal interpretations is not possible). His argument goes as follows:

(1) Let θ be a sentence that is true in just those models in which neither ϕ nor $\sim\phi$ is true.

Footnote 4 (continued)
framework of propositional logic, e.g. by allowing multiple-conclusions [like Carnap (1943) and Shoesmith and Smiley (1978) suggested], or by resorting to a bilateral formalisation of logic [see e.g. Smiley (1996) and Rumfitt (2000)], categoricity can be regained.

[5] One moral that we can drawn from Carnap's discovery of the non-normal interpretations is that syntactical uniqueness is not a sufficient condition for semantic uniqueness, i.e., for determining a unique meaning for the syntactical connectives.

(2) There are no models in which θ and ϕ are both true. (1)
(3) $\{\theta, \phi\} \vdash \curlywedge$ (2)
(4) $\{\theta\} \vdash {\sim}\phi$ (3) (Rule: If $\Gamma \cup \{\phi\} \vdash \curlywedge$, then $\Gamma \vdash {\sim}\phi$)
(5) There are no models in which θ and ${\sim}\phi$ are both true. (1)
(6) $\{\theta, {\sim}\phi\} \vdash \curlywedge$ (5)
(7) $\{\theta\} \vdash {\sim}{\sim}\phi$ (6) (Rule: If $\Gamma \cup \{\phi\} \vdash \curlywedge$, then $\Gamma \vdash {\sim}\phi$)
(8) $\{{\sim}\phi, {\sim}{\sim}\phi\} \vdash \curlywedge$
(9) $\{\theta\} \vdash \curlywedge$ (4), (7), (8)
(10) There are no models in which θ is true. (9)
(11) $\curlywedge$ (1), (10)
(12) In every model, either ϕ is true or ${\sim}\phi$ is true. (1), (11)

The general structure of the argument could be seen as a *reductio ad absurdum*.[6] It is assumed that the language of propositional logic is extended by adding a sentence that is true *just* in those models in which neither ϕ nor ${\sim}\phi$ is true. This is an instance of McGee's (2000: p. 70) general assumption that "for any class of models, there is a mathematically possible language in which there is a sentence true just in those models." On this assumption, by validly reasoning in the meta-theory, it follows (at line 9) that θ is inconsistent and, thus, that it has no model (at line 10). Therefore, McGee concludes that in every model, either ϕ is true or ${\sim}\phi$ is true (at line 12).

It seems to me that the derivation of (12) is, in a broader sense, a *non-sequitur*.[7] What we could legitimately do after line (11), when we find out that there is a contradiction between (1) and (10), is to deny assumption (1), i.e., to derive that: (11`) it is not the case that $\theta$ is a sentence that is true just in those models in which neither $\phi$ nor ${\sim}\phi$ is true. However, (11`) could be true in two cases: (a) θ is a sentence that is true in those models in which neither ϕ nor ${\sim}\phi$ is true, but not only in them, or (b) there are no models in which neither ϕ nor ${\sim}\phi$ is true and, thus, *a fortiori*, θ could not be true. McGee offers no reason for excluding option (a). However, due to Carnap's results, I argue that (a) is what actually makes (11`) true.

[6] The argument starts with the assumption that θ is a sentence that *is true* in just those models in which neither ϕ nor ${\sim}\phi$ is true. Then, by valid reasoning we find out, at line (9), that θ is inconsistent, i.e., it has no model (10). Since (1) and (10) constitute an inconsistent pair of sentences -(1) says that θ has at least one model and (10) says that θ has no models-, an absurdity follows at line (11). What led us, however, from the very beginning to this inconsistency was the assumption (1), which does not simply say that θ is true in general, but that it is true *just* in a certain class of models. Therefore, the negation of (1) has to be inferred by *reductio* and not simply the negation of θ. Actually, as Carnap showed, if θ is true in those models in which neither ϕ nor ${\sim}\phi$ is true, then θ is true in all the models of PC.

[7] As a reviewer kindly suggested, someone may say that since assumption (1) leads to a contradiction, then one may classically draw any conclusion whatsoever from this contradiction. Hence, the derivation of (12) from it is not a non-sequitur. Indeed, from a strictly formal point of view, it is not a mere non-sequitur. However, when we arrive at a contradiction, it is *more reasonable* to see what false ideas led us to that contradiction, and not to start deriving any conclusion from it. Since McGee's argument is in the meta-theory, and uses both proof-theoretic and model-theoretic resources, and it is a logical fact (more precisely, a model-theoretic fact) that there are models in which a sentence and its negation are both false, the truth value of the starting assumption should be first investigated.

The propositional calculus is sound with respect to a model in which neither ϕ nor ~ ϕ is true and, as a matter of mathematical fact, there exists such a model, e.g., the one that satisfies each and every theorem of the calculus and satisfies no non-theorem (let us note with N this model). To see that the propositional calculus is sound in this model, let Γ be an arbitrary set of premises and σ an arbitrary sentence in the language of propositional calculus, and let us further suppose that $\Gamma \vdash \sigma$. There are two cases to be considered: Γ contains only theorems or it contains at least one non-theorem. If Γ contains only theorems, then σ will also be a theorem and, thus, true in the model N. If Γ contains at least one non-theorem, then the sequent $\Gamma \vdash \sigma$ will be valid even if σ is false.[8]

McGee assumes that θ is true in just those models in which neither ϕ nor ~ ϕ is true (let M be this class of models). Since neither ϕ nor ~ ϕ is true in N, it follows that N is a member of M. Thus, due to assumption (1), θ is true in N. But if θ is true in N, it follows that θ is a theorem, because only the theorems are true in N. However, if θ is a theorem of the propositional calculus, it cannot be true *just* in M, but in all the models of the propositional calculus. Thus, the starting assumption of McGee is false; θ is not true *just* in M. We see thus that the existence of the model N falsifies assumption (1).

Now, since we have a logical reason to take assumption (1) to be false, we may reconsider McGee's argument. Naturally, since assumption (1) is false, it leads to a contradiction. Formulated explicitly, assumption (1) is a conjunction of "θ is true in M" and "there are no other models, besides those from M, in which θ is true". As we have noticed, since assumption (1) leads to a contradiction, we have to deny this assumption. Hence, by DeMorgan's rules, we obtain the disjunction of (A) "it is not the case that θ is a true sentence in M" and (B) "there is at least one model, besides M, that satisfies θ". This disjunction is indeed true, because θ could be either a theorem, or a non-theorem. If θ is a theorem, then (B) is true, because all the theorems of propositional calculus are true in all the models of propositional calculus, not only in M, and, thus, (11`) is true. If θ is not a theorem, then (A) is true, because all non-theorems are false in N and, since N belongs to M, θ will not be generally true in M. Hence, (11`) is true. Therefore, McGee's argument from (1) to (11) is valid, but the derivation of (12) is, in a broader sense (see footnote 7), a *non-sequitur*.

Another way of looking to McGee's argument is to use the resources of set theory. We can read McGee's argument as starting from the universal sentence (1) $\ulcorner(\forall W)[(W \vDash \theta) \leftrightarrow ((W \nvDash \varphi) \,\&\, (W \nvDash {\sim}\varphi))]\urcorner$, where W is an arbitrary model. We should note that from this statement, by the distribution of the universal quantifier over implication, we get something to the effect that (2) $\ulcorner(\forall W)((W \nvDash \varphi) \,\&\, (W \nvDash {\sim}\varphi)) \rightarrow (\forall W)(W \vDash \theta)\urcorner$ which, with the standard reading of the quantifiers implies (3) $\ulcorner(\forall W)((W \nvDash \varphi) \,\&\, (W \nvDash {\sim}\varphi)) \rightarrow (\exists W)(W \vDash \theta)\urcorner$. McGee's argument stated above, however, shows that the consequent of this statement is inconsistent, i.e., (4) $\ulcorner(\exists W)(W \vDash \theta) \vdash \curlywedge\urcorner$, and thus it is not true. Hence, from (3) and (4), by contraposition, it follows that (5) $\ulcorner{\sim}(\forall W)((W \nvDash \varphi) \,\&\, (W \nvDash {\sim}\varphi))\urcorner$. By DeMorgan rules, what follows from (5) is an existential statement, i.e., (6) $\ulcorner(\exists \mathrm{W})((W \vDash \varphi) \vee (W \vDash {\sim}\varphi))\urcorner$, and

8 This proof could be found in a different terminology in Carnap (1943: pp. 91–92).

not an universal one, as McGee would want. This last statement, however, is perfectly compatible with the existence of the non-normal interpretation of the second kind for classical negation.

As McGee (2000: p. 72) puts it, correctly I think, we have to go beyond language in order to determine whether a sentence is true in a particular model (so to say: "we have to take a look at the model"). We cannot simply stipulate that in any case there is a sentence that is true *just* in a certain class of models.[9] Actually, since the soundness theorem for classical propositional logic is insufficient to pin down uniquely the intended meanings of all its connectives, the open-endedness understanding of the rules will neither work for this job. An open-ended understanding of the rules for negation will not be able to eliminate a model in which a sentence and its negation are both false and their disjunction is true. The propositional calculus is sound and complete with respect to this model. Thus, an extension of the propositional language with a sentence that is true *just* in the class of models in which a sentence and its negation are false is not a "mathematically possible extension". The existence of the non-normal interpretations of the second kind for classical propositional calculi shows that McGee's (2000: p. 70) assumption that "for any class of models, there is a mathematically possible language in which there is a sentence true just those models" is not universally true.

3 Two objections and replies

3.1 Objection 1

The propositional calculus (PC) allows for non-normal interpretations in which all theorems are true and all non-theorems are false. In these interpretations the syntactical sign for negation has a non-normal meaning. McGee claims that this interpretation can be ruled out if we allow an expansion of the initial language L in which a new sentence θ is true in all and only these non-normal interpretations. My claim above was that if θ is true in a non-normal interpretation, then θ must be a theorem of PC and hence θ must be true in other interpretations as well, contrary to its definition. However, the criticism states, this seems to overlook the fact that θ is in an expansion of L, not in L. Thus, it does not follow that θ is a theorem of PC. Consequently, if one distinguishes the initial language of PC from its expansion, the definition of θ appears to be consistent.

[9] As a reviewer suggested to me, the main point made in this paper can be formulated by saying that McGee's argument already presupposes the notion of an admissible model, i.e., a model that respects the meanings of the logical constants. For if we let McGee's claim be about any model, we get that the claim must also hold for non-standard interpretations of propositional logic, e.g. interpretations that make both φ and not-φ true. But there is no sentence that is (actually) true in exactly those models.

3.2 Reply to objection 1

The problem is whether L can be consistently expanded with a sentence that is true *just* in the models in which a sentence and its negation are both false. The objection states that if θ is in the expansion of L and not in L, then it does not follow that θ is a theorem, and thus true in all the models of PC. If we consider this problem carefully, however, if PC allows for non-normal interpretations in which a sentence and its negation are both false when it is formulated in the extended language, then θ *must* be a theorem. This is a *logical result* for which there is a rigorous proof given by Carnap (1943: pp. 91–92), according to which the only sentences that are true in the models in which a sentence and its negation are both false are the theorems of PC. Thus, the definition of θ, as being true *just* in the class of models in which a sentence and its negation are false, is problematic. There is no such θ. Being a theorem, θ is true in all models of PC. To reiterate, if θ is to be true in the models in which a sentence and its negation are false, it *must* be a theorem.

3.3 Objection 2

The point of the paper is that (12) is *demonstrably false* and thus McGee's argument should be regarded not as a proof of (12), but as a rejection of (1). The problem with taking the argument as a rejection of (1) lies in the way in which the rules of PC are formulated. Carnap's proof uses a system of PC taken from Hilbert and Bernays, in which all the axioms are logically valid, and the rules of inference (substitution and modus ponens) are validity preserving. But McGee uses a natural deduction system in which not all the rules are validity preserving. In particular, *reductio ad absurdum* is not validity preserving. If the sentence ψ is neither valid nor contradictory, the inference from ψ to falsehood is not validity preserving. Thus, the criticism in the paper does not apply to McGee's system.

3.4 Reply to objection 2

It is well known that the natural deduction formalization of propositional logic and the axiomatic formalizations are equivalent, in the sense that if a formula is deducible from a set of formulas by the natural deductive rules for the propositional connectives, then it is also deducible by the rules and axioms of the axiomatic systems (and the other way around). In this sense, Carnap's results obtained on an axiomatic form of PC remain sound even if a natural deduction system is considered. As a matter of fact, Carnap (1943: p. 8) was aware of the fact that 'the different forms (of PC) vary with respect to the choice of primitive signs, primitive sentences, and rules of inference, but they are known to agree with respect to possible results of proofs and derivations'. Actually, the syntactic results are formulated by Carnap (1943: pp. 16–19) in a general manner, not to pure forms of PC, but for calculi containing PC, as they are usually used in the logical foundations of deductive sciences. On the basis of these general syntactic formulations, in the case of the non-normal interpretations “the results hold likewise for any other form of PC” (Carnap 1943: 69), not only for the system taken from Hilbert and Bernays.[10] Therefore, the problem whether *reductio* is validity preserving is not relevant in this case; (12) remains demonstrably false.

Acknowledgements I would like to thank my audiences in Hejnice, Klagenfurt, Novara, and Tübingen, where parts of my work on this paper have been presented, in particular Andrzej Indrzejczak, David Miller, Jaroslav Peregrin, Andrea Sereni and Peter Schroeder-Heister, for helpful discussions. Special thanks to Gabriel Sandu, Iulian Toader, and the reviewers for their useful feedback.

References

Belnap, N. D. (1962). Tonk, plonk and plink. *Analysis, 22,* 130–134.

Bonnay, D., & Westerståhl, D. (2016). Compositionality solves Carnap’s problem. *Erkenntnis, 81*(4), 721–739.

Carnap, R. (1943). *Formalization of logic*. Cambridge, MA: Harvard University Press.

Došen, K., & Schroeder-Heister, P. (1985). Conservativeness and uniqueness. *Theoria, 51*(3), 159–173.

Harris, J. H. (1982). What’s so logical about the ‘logical’ axioms? *Studia Logica, 41,* 159–171.

Hjortland, O. T. (2014). Speech acts, categoricity and the meaning of logical connectives. *Notre Dame Journal of Formal Logic, 55*(4), 445–467.

Humberstone, L. (2011). *The connectives*. Cambridge: MIT Press.

Kalicki, J. (1950). Note on truth-tables. *The Journal of Symbolic Logic, 15*(3), 174–181.

McGee, V. (2000). Everything. In G. Sher & R. Tieszen (Eds.), *Between logic and intuition*. Cambridge: Cambridge University Press.

McGee, V. (2015). The categoricity of logic. In C. R. Caret & O. T. Hjortland (Eds.), *Foundations of logical consequence*. Oxford: Oxford University Press.

Rumfitt, I. (2000). ‘Yes’ and ‘no’. *Mind, 109,* 781–823.

Scott, D. (1971). On engendering an illusion of understanding. *The Journal of Philosophy, 68*(21), 787–807.

Shoesmith, D. J., & Smiley, T. J. (1978). *Multiple-conclusion logic*. Cambridge: Cambridge University Press.

Smiley, T. (1996). Rejection. *Analysis, 56*(1), 1–9.

Tarski, A. (1938). Der Aussagenkalkül und die Topolagie. *Fundamenta mathematicae, 31,* 103–134.

[10] Certainly, Carnap refers to the standard formulations, i.e, those with a single conclusion. If we use multiple conclusions, categoricity can be regained. (see Hjortland 2014).