Is modus ponens a valid inference rule?*

Rosen Lutskanov

Abstract. The present text is organized as follows: (§1) examines briefly the role that modus ponens plays in traditional logic, (§2) discusses four well-known (alleged) counterexamples that cast doubt on its validity, (§3) analyses the interpretation of these counterexamples, as it was proposed by their inventors, (§4) marks the common features of the possible strategies to overcome the criticisms against modus ponens, (§5) considers the conclusions that can be based on the proposed analysis.

Keywords: logical consequence, inference, conditionals, modus ponens, logical form

§ 1. Presently in the theoretical domain of logic almost nothing seems certain. One common point of departure is that it has changed dramatically over the last hundred years. The numerous crises and revolutions in its recent history have removed the classical logic from the privileged position where it was instituted by Boole, Frege, Peirce, Russell and Hilbert. The introduction of many non-classical logics was motivated in many different ways - changes in the proposed semantics, applications of logic in different fields, and philosophical analyses of natural language. Usually these shifts correspond to criticisms against some of the classically valid logical laws and rules of inference.¹ For example, there are

---

* I am grateful to the two anonymous referees of this paper, who provided extended comments on a previous version of this paper. I have done my best to incorporate their remarks in the text, bearing in mind the limited space and the original design of this paper.

¹ A suggestive example is intuitionism. The originator of this school of thought, Brouwer, expressed his suspicions concerning the application of some classically valid inference rules in mathematical proofs. Later his pupil Heyting (and independently Kolmogorov) have devised a new semantics which corresponds (at least to some extent) to the intuitions of Brouwer (in order to stress this fact, it was named BHK interpretation). Later some philosophers, notably Michael Dummett, applied this semantic approach in order to solve some typically philosophical problems. I am grateful to one of the anonymous referees, who has drawn my attention to the connection between these developments.
lots of well-known criticisms against the Law of Excluded Middle (according to which, for a pair of contradictory propositions, p and ~p, at least one of them is true) or the rule of reductio ad absurdum (according to which the derivation of contradiction from p licenses the derivation of ~p). Other laws and rules of inference are virtually never contested - the reason is that they seem too deeply entrenched in our intuitions about the meaning of the word “logic”. In particular, there is a rule of inference which seems completely sheltered from justified criticism. This is the so-called modus ponens (known also under other names - modus ponendo ponens, affirming the antecedent, rule for implication elimination, rule of detachment). This inference rule was known and applied already in Antiquity\(^2\) and is intimately connected with other rules of ancient origin - modus tollens, hypothetical syllogism and the constructive dilemma. Its form is p, (p \(\rightarrow\) q) \(\setminus\) q, which means that q is inferable from p and the material implication with antecedent p and consequent q. This inference can be justified purely semantically: if p is true, then the only way for p \(\rightarrow\) q to be also true is q to be true, since any material implication with true antecedent and false consequent is false. That is why, this rule of inference (which is formulated on the level of meta-language), corresponds to the tautology p \&(p \(\rightarrow\) q) \(\rightarrow\) q.

Modus ponens can be seen as part of the definition of the notion of material implication, as implicitly determining the meaning of the construction “if ..., then ...”. Should it be seen then as immune to revision? There are at least four notable attempts to show that modus ponens is invalid.\(^3\) While we discuss them, it is important to bear in mind that our task shall be to see whether it is valid not just for material implication, but for the indicative conditionals in general. By modifying the counterexamples a bit, they can target also the subjunctive conditionals, but this is another story (Bledin 2015, §1). Here I am going to examine these counterexamples and identify their common features.\(^4\)

§ 2.1. The first criticism against modus ponens that we are going to examine is formulated by Jordan Howard Sobel in 1970 and is documented by David Lewis in his book Counterfactuals (1973). It is stated by means of sequences of propositions of the following form:

\(^{2}\) Although it cannot be located in the writings of Aristotle, a similar scheme can be found in the Stoics. It is certain that the inferences of this type were identified as valid by virtue of its logical form around 2nd century AD at the latest (Bobzien 2002).

\(^{3}\) There are many other similar attempts which I am not going to discuss here, since in my opinion they are much less convincing, cf. for example Willer 2010 and Kolodny, MacFarlane 2010. Modus ponens is attacked also in the framework of paraconsistent logic, the logics of vagueness, and many different doxastic and deontic formalisms. Their peculiarities are beyond the scope of the present paper.

\(^{4}\) One of the anonymous referees has suggested that the alleged counterexamples propose an inadequate formalization of the natural language statements that they are employing and would be better to look for another logical framework, for example dynamical logic. There are good reasons backing this claim (at least for some of the counterexamples that we are going to discuss below). My task here is to present these alleged counterexamples in their original form, and then to try to find out what is their meaning, whether they are real “refutations” of modus ponens, and what is their relevance to the question of the nature of logical consequence.
(1) If Albert comes to the party, then it would be great.
(2) If Albert and Betty come to the party, then it would be awful.
(3) If Albert, Betty, and Carl come to the party, then it would be great, etc.

The sequence can be prolonged indefinitely but this is not necessary since we have already seen one of its important features - the fluctuation of the consequent - it ascribes to the party the alternating and incompatible properties to be “great” and “awful”, which means that the conditionals in the sequence have incompatible consequents. Let us now grant that there are reasons to believe in the truth of the first two propositions in the sequence. Imagine for example that Albert is known as a witty conversationalist (which justifies the acceptance of (1)) but he has recently broken his relationship with Betty, which means that he shall be very moody in her presence (which justifies the acceptance of (2)). Thus, we immediately get an unacceptable conclusion which can be reached by means of modus ponens and correspondingly casts doubt on this rule of inference. Indeed, let us examine the first two members of the sequence and moreover assume that Albert and Betty both came to the party. Then from (1) by modus ponens we come to the conclusion that the party is going to be great and from (2), again by modus ponens, we come to the conclusion that the party is going to be awful. As far as the party cannot be great and awful in one and the same time, this inference is unacceptable (Lycan 1993, 415). If we set $A = \text{“Albert comes to the party”}$, $B = \text{“Betty comes to the party”}$, $C = \text{“The party would be great”}$, $D = \text{“The party would be awful”}$, then (1) becomes $A \rightarrow C$, and (2) becomes $A \& B \rightarrow D$. Correspondingly, the argument can be formalized in the following way:

(i) $A \& B$ (we assume that both Albert and Betty are coming to the party)
(ii) $A$ (from (i) by conjunction elimination)
(iii) $A \rightarrow C$ (premise 1)
(iv) $C$ (from (ii) and (iii) by MP)
(v) $A \& B \rightarrow D$ (premise 2)
(vi) $D$ (from (i) and (v) by MP),

where (iv) and (vi) are incompatible. As is easily seen, the unacceptable conclusion is reached by means of modus ponens and the seemingly unproblematic rule for conjunction elimination which states that if we infer a conjunction then we can infer also each of the conjuncts alone.

§ 2.2. A second counterexample is proposed by Allan Gibbard and concerns the seemingly harmless statement “If you insult me, I will be kind, but if you insult my wife, I won’t be kind”. Let us set $A = \text{“You insult me”}$, $B = \text{“You insult my wife”}$, $C = \text{“I will be kind”}$. Then the statement shall have the form $(A \rightarrow C) \& (B \rightarrow \neg C)$. Let us imagine that I am extremely rude - I insult both Gibbard and his wife (A&B). Thus, we get the following inference:

(i) $A \& B$ (I insult both of them)
(ii) $(A \rightarrow C) \& (B \rightarrow \neg C)$ (premise)
(iii) $A$ (from (i) by conjunction elimination)
(iv) $A \rightarrow C$ (from (ii) by conjunction elimination)
(v) $C$ (from (iii) and (iv) by MP)
(vi) B (from (i) by conjunction elimination)
(vii) B → ~C (from (ii) by conjunction elimination)
(viii) ~C (from (vi) and (vii) by MP),
where (v) and (viii) are contradictory (Lycan 1993, 420). Here, just as in the first case, the only additional rule of inference we need to employ is conjunction elimination.

§ 2.3. Our third counterexample was proposed by Vann McGee in a famous paper published in 1985 (McGee 1985). He asks us to consider the following case. In USA the Presidential elections are forthcoming and the nominees are Ronald Reagan and John Anderson for the Republicans and Jimmy Carter for the Democrats. According to the polls, Reagan and Carter are way ahead of the third candidate and Reagan is an almost certain winner. In this case it is seemingly reasonable to assume that (1) “If a Republican wins the election, then if it’s not Reagan who wins it will be Anderson” and (2) “A Republican will win the election”, although it seems that we are not justified to assume that (3) “If it’s not Reagan who wins it will be Anderson” (McGee 1985, 462). On the contrary, the data suggests that (4) “If the winner is not Reagan, then it is Carter”, since the chances of Anderson are meager. It seems that we should grant (1) since Reagan and Anderson are the only candidates from the Republican Party. Moreover, it seems that we should grant (2), since the polls suggest that Reagan is ahead of Carter. Finally, we should not grant (3), since Carter is ahead of Anderson. Let us then set A = “Reagan will win the election”, B = “Anderson will win the election”, C = “Carter will win the election”. Then we get the following argument:

(i) A V B → (~A → B) (premise 1)
(ii) A V B (premise 2)
(iii) ~A → B (from (i) and (ii) by MP)
(iv) ~A → C (assumption 4).

Thus, we get a paradoxical conclusion which was left unnoticed by almost all commentators: since B and C are incompatible (Anderson and Carter cannot both win the elections), then it follows that it is impossible for Reagan to loose, i.e., he shall win the elections by logical necessity! Indeed, if Reagan loses (~A), then we get B (from (iii) by MP) and C (from (iv) by MP), which as we have already seen are incompatible.

§ 2.4. The last counterexample that we are going to examine here was formulated by William Lycan. He invites us to imagine that X attends a meeting of a radical leftist group, although his political leanings are more or less conservative. In fact, he went to the meeting since a friend of his asked him to do so. During the meeting some people have shared their suspicions that there is CIA agent who was infiltrated in the group. X does not believe that this is the case but assumes that if this is the case, then he could have problems (he believes that if the meeting is under the surveillance of CIA, then his name shall be blacklisted). What X (thinks he) knows is a sufficient reason to accept this implication. What X does not know is that there is in fact a CIA agent present,
who knows that X supports the government, so nothing bad is going to happen
to him (Lycan 1993, 414). The propositions (1) “If there is a CIA agent present
at the meeting, X is going to have problems”, (2) “There is a CIA agent present
at the meeting” and (3) “X is not going to have problems” are inconsistent if
modus ponens is available. Indeed, let us label (2) by A and (3) by ~B, which
means that (1) is (A → B). From (1) and (2) we get B (by MP) which contradicts
(3). What do these counterexamples show? Are we forced to sacrifice modus
ponens? Is there another way out?

§ 3. What is the meaning of these counterexamples according to their
inventors? According to McGee his example shows that modus ponens “is not
strictly valid; there are occasions on which one has good grounds for believing
the premises of an application of modus ponens but yet one is not justified in
accepting the conclusion” (McGee 1985, 462). This manifests his conviction that
the inference rules of logic are not purely semantical: they don’t show just what
propositions follow from a particular set of premises but also what is the relation
between the reasons to accept the premises as true and the reasons to grant the
conclusion as well. This is just as it should be, since logic demarcates the minimal
conditions which rational thinking in general must meet. Furthermore, in
his opinion in this case we have been misled by a well-known fallacy - hasty
generalization. We have considered many inferences of this form, where the
acceptance of the premises entails that we are prepared to accept the conclusion
as well. On this basis we have reached a seemingly correct but unjustified
conclusion - that modus ponens is a formally valid rule of inference. In fact,
it is not valid, since it does not behave properly for one particular syntactically
characterized class of propositions, i.e., propositions where the consequent of the
implication itself has the form of an implication (McGee 1985, 464-465, 468).
Otherwise put, modus ponens seems sacrosanct because usually or in virtually
all cases we apply it to propositions with a relatively simple structure.

§ 4. We can safely assume that this is not the end of the story, that many
vociferous defenders of modus ponens shall come to the rescue. What weakness
are they able to spot in the arguments we have examined above? There are
four possible types of counter-attack: (A) these are not real counterexamples,
since the conclusions are false or unjustified; (B) the counterexamples do not
display correctly the logical form of the premises and the conclusion; (C) the
counterexamples are correct but do not affect modus ponens - there is another
inference rule which is to be blamed; (D) the counterexamples do not show what
they are aiming to show, since at least one of the premises in each one of them
is in fact false.

§ 4.1. Let us start with the first type of response. It is easiest to see how it
works if we go back to the third counterexample. Let us reiterate the main point:
the information that we are presented with allows us to claim with sufficient
confidence that the winner in the presidential elections shall be Reagan (it is not
at all easy to abstract from the fact that this has in fact happened). Accordingly,
the conclusion that we claim is not justified (if Reagan loses, then Anderson
is going to win) and the conclusion that we claim is justified (if Reagan loses, then Carter is going to win), are trivially true, since they have the form of an implication with a false antecedent. That is why there is no reason to prefer one of them at the expense of the other, or to claim that one of them is incorrect, while the other is not (Fulda 2010, 272). This argument is not very convincing: it provides an effective defense of modus ponens, but the price is too high - to block the choice between the incompatible alternatives.

§ 4.2. The second type of criticism targets Sobel’s counterexample, although it is relatively easy to modify the argument to be presented below for the Gibbard case. As we have already seen, the first two propositions in the sequence have the form $A \rightarrow C$ and $A&B \rightarrow D$. This suggests that the problem is generated by the classically valid rule of strengthening the antecedent, which says that if we are given an arbitrary proposition $A \rightarrow C$, then for each other proposition $X$ we can infer $A&X \rightarrow C$. It is easy to justify this rule semantically: if $X$ is true, then the truth of $A&X \rightarrow C$ is entailed by the truth of $A \rightarrow C$, if $X$ is false, then $A&X \rightarrow C$ is trivially true, since its antecedent is false. Although semantically innocent, the rule cannot be defended from an epistemic point of view. It is natural to assume that the inflow of new information can affect the acceptability of previously established claims. That is why the acceptance of conditionals of the form $A \rightarrow C$ is justified just in case we did not accept or did not render as probable any proposition $X$ which is able to annihilate the connection between $A$ and $C$. In our case $B$ is precisely such proposition, which makes the transition from $A$ to $C$ impossible. This suggests that the proposition $A \rightarrow C$ is in fact implicitly elliptical and has the form “$A$ implies $C$, if we exclude $B$”, i.e., $A&\neg B \rightarrow C$. In this way we can accept both propositions - their antecedents shall be mutually incompatible, so it would not be puzzling at all that their consequents are likewise incompatible. Lycan objects to this move: if we assume that the antecedents of all conditions are elliptical, then it would be virtually impossible to account for their truth conditions (Lycan 1993, 419-420). The problem is that the additional conditions cannot be determined on the basis of the conditional, taken in isolation.

§ 4.3. A third possible strategy is to claim that in fact the counterexample does not affect modus ponens. To see how this works, let us reconsider McGee’s counterexample, paying attention to the grammatical form of the corresponding sentences. The first one was: “If a Republican wins the election, then if it’s not Reagan who wins it will be Anderson”, and the third was “If it’s not Reagan who wins, it will be Anderson”. If we compare their consequents, we shall see that, although they are seemingly similar, their grammatical form is different. The reason is that “straightforwardly embedding one such conditional within another as its consequent has the effect of transforming the tense of the embedded conditional from the future to the future future, so that the embedded conditional is not in fact equivalent to a non-embedded conditional of the same form” (Lowe 1987, 47). This remark seems to be justified but it is hopeless to discard all alleged counterexamples by means of the same strategy, since they do not employ this embedding technique.
§ 4.4. A forth possible way out is to claim that some of the premises are in fact false. In this case the problem disappears since a valid inference with a false premise is not guaranteed to have true conclusion. Let us return to Sobel’s case and assume that Albert comes to the party (this is a minimal requirement for the truth of the antecedents of the conditionals in the sequence, since if Albert does not come the antecedent becomes false and correspondingly the conditionals would be trivially true). In this case four possibilities are left open: (a) Betty comes and the party is great; (b) Betty comes and the party is awful; (c) Betty does not come and the party is awful; (d) Betty does not come and the party is great. Now let us consider them: if (a) is the case, then the second member of the sequence becomes false, since it claims the opposite, if (b) is the case, then the first member of the sequence becomes false, since it shall have true antecedent and false consequent, if (c) is the case, then the first one is false again, if (d) is the case, then both are true but the second is trivially true, since it has a false antecedent and a false consequent. This means that Sobel’s example shows just that the only way to keep both premises true is (d) which is not a problem for modus ponens as such (Sinnott-Armstrong 1999, 126-127).

We can say the same about Lycan’s case: here we have again an implication with affirmed antecedent and denied consequent, which is therefore false. This is going to work also for McGee’s case - it shows just that the first premise cannot be true, if it is treated as a material implication, since a material implication with true antecedent (affirmed in the second premise) and false consequent (denied by rejecting the conclusion) is false by definition. Therefore, either the conditional is not interpretable as material implication, which means that it is irrelevant to the question of the validity of modus ponens, as it is traditionally construed, or that it does not show what McGee claims it is showing. In either case his attack against modus ponens fails (Katz 1999, 412). Another argument is aimed against the second premise. The first premise in McGee’s inference has as its antecedent the claim that a Republican is going to win. This assumption is affirmed in the second premise, which makes it relevant for the argument as a whole. Therefore, either the second premise is false, or it is not true that if Reagan would lose, then Carter is going to win, which seems probable in view of the polls (Over 1987, 143).

§ 5. What conclusion can be made on the basis of this extended discussion? On the one hand, there is the possibility to admit our defeat and claim that modus ponens is in fact not valid. On the other hand, we can apply some of the representative “conventionalist stratagems” (in the sense of Lakatos), which we have examined above - to formulate an ad hoc refutation of the alleged refutation. There is also a third possibility - to assume that there is a real problem and it is much deeper, since it is concerned with the very heart of logic, with the question about its principal subject matter. In the last century three different answers of this question were proposed: it was claimed that logic is primarily concerned with (1) truth (semantic conception), (2) justification (epistemic conception), (3) information (information-theoretic conception). Each of these answers has significant impact on our rendering of validity - as preservation of truth, justification or information content. (1) was most clearly exemplified in
the approach of Tarski and the now standard definition of validity. (2) is related to the work of Frege and Brouwer, who were not primarily concerned with truth and its preservation, but with our reasons to accept something as true. (3) is relatively new and is exemplified in the work of Seth Yalcin on the nature of epistemic modalities. As we are going to see, it is relevant to the discussion of the problems we have sketched above. According to Justin Bledin (Bledin 2015, §3) these problems stem from our attempt to link (1) and (2) by means of a bridge principle of the following type: If an argument form is logically valid and we have good reasons to accept each of its premises, then we have good reasons to accept its conclusion. In McGee’s counterexample we have good reasons to accept the premises, but not the conclusion, which forces on him the disputable doctrine that modus ponens is not valid. The same applies to the other counterexamples: it seems that we have good reasons to accept the claim that the party is going to be great and that the party is going to be awful (case 1), that I am going to be kind and rude (case 2), that I shall have problems with the authorities and everything is going to be fine (case 4). This leads Bledin to the radical conclusion that logical inference is not at all about truth, justification, or the link between them, but about information content, i.e., about “the preservation of structural properties of bodies of information that we generate, encounter, absorb, and exchange as we interact with one another and learn about our world” (Bledin 2015, §3). According to this understanding of logic, an implication means that if we apply a minimal modification of the data at our disposal so that the information contained in the antecedent is incorporated in our present information state, then the information contained in the consequent shall be also incorporated (of course, this definition presupposes a non-trivial idealization: that we can specify uniquely the minimal modification of the available data which is compatible with the antecedent). Let us now analyse in turn the four alleged counterexamples we have examined above and see what follows for them in view of the suggested conception.

§ 5.1. Let us start with Sobel’s case. A minimal modification of the available data can incorporate either the claim that Albert is coming alone, or that Albert is coming and Betty is coming (although the scenario does not specify what exactly is going to happen). This means that we can affirm either the first or the second proposition in the sequence, but not both. We can reason in the same way in relation to the second and the third proposition. This blocks the derivation of a contradiction - the real problem is in the premises, since they cannot be affirmed simultaneously (this solution comes close to the fourth strategy that we have sketched above, but does not coincide with it since it provides us with a general method, exposing the incompatibility of the premises in all alleged counterexamples).

§ 5.2. Let us now consider the Gibbard’s case. It was assumed that I was rude with both him and his wife. The subsequent events would provide information which is incompatible either with the first proposition in the conjunction, or with the second, although we cannot say now which one needs to be rejected. In either case the conjunction itself would be false. This means that even if we
are not able to determine the minimal modification in our information state, we can be sure that the premise of the inference is false and there is no reason to blame modus ponens.

§ 5.3. In McGee’s case, if we take into account both the data from the polls and the subsequent events, we can say that the minimal modification would incorporate the assumption that Reagan is going to win. This blocks the affirmation of all claims which feature the negation of this proposition in the antecedent of a conditional (there is no available information modification which incorporates it in our information state). In particular, the first premise turns out to be false.

§ 5.4. In Lycan’s example we can envisage different scenarios - for example, after the meeting the secret agent can establish a contact with X and tell him that he is aware of his pro-government attitudes. This can be seen as a minimal revision of the available data, which provides the information that the CIA agent was in fact present. In this case the problem shall be again with one of the premises.

§ 5.5. Thus, we have seen that if we follow the footsteps of Bledin and interpret validity as unrestricted informational incorporation preservation, then our problems are bound to disappear. Do the examined counterexamples to modus ponens really force on us a radical change of the received views about validity by adding a new parameter (which corresponds to information content or information state)? The answer of this question depends on our assessment of the proposed solution. Is it (a) effective, (b) motivated, (c) not unjustifiably radical, and (d) able to become the foundation of a new logical paradigm? As far as (a), (b), and (c) are concerned, I believe that we can reach a more or less unanimous verdict. As far as (d) is concerned, many authors have proclaimed the coming of a new era - the so-called “informational turn”. The works of Barwise on situation semantics, of van Benthem on the dynamics of information, the new epistemic logics and the so-called inquisitive semantics can be seen as parallel developments which correspond to this understanding of the nature of logic. The future is going to show what is its heuristic potential. Even now its potential to tackle the alleged counterexamples against modus ponens speaks in its favour.

References


Assoc. Prof. Rosen Lutskanov, PhD
Institute for the Study of Societies and Knowledge
Bulgarian Academy of Sciences
4 Serdika Str.
1000 Sofia, Bulgaria
Email: rosen.lutskanov@gmail.com