Università degli Studi di Milano

Modus ponens and *ought*

Sandro Zucchi

2022-2023

The miners are trapped

Kolodny and MacFarlane (2010)

> *The miners are trapped either in shaft A or in shaft B, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If we block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.*

Action	*if miners in A*	*if miners in B*
Block shaft A	*All saved*	*All drowned*
Block shaft B	*All drowned*	*All saved*
Block neither shaft	*One drowned*	*One drowned*

What to do?

Kolodny and MacFarlane (2010)

We take it as obvious that the outcome of our deliberation should be

(1) We ought to block neither shaft

Still, in deliberating about what to do, it seems natural to accept:

(2) If the miners are in shaft A, we ought to block shaft A.

(3) If the miners are in shaft B, we ought to block shaft B.

We also accept:

(4) Either the miners are in shaft A or they are in shaft B.

But (2), (3), and (4) seem to entail

(5) Either we ought to block shaft A or we ought to block shaft B.

And this is incompatible with (1). So we have a paradox.

The argument

formal reconstruction

Kolodny and MacFarlane propose the following formal reconstruction of the argument:

1	$inA \vee inB$	
2	if inA, $O(blA)$	
3	if inB, $O(blB)$	
4	inA	
5	$O(blA)$	2, 4, MP
6	$O(blA) \vee O(blB)$	5, $\vee$ intro
7	inB	
8	$O(blB)$	3, 7, MP
9	$O(blA) \vee O(blB)$	8, $\vee$ intro
10	$O(blA) \vee O(blB)$	1–9, $\vee$ elim

Montague-Kalish-Gettier-style deduction

- Here is a version of the proof in the deduction system we are familiar with:

1.	$inA \vee inB$	P
2.	$if\ inA,\ O(bIA)$	P
3.	$if\ inB,\ O(bIB)$	P
4.	~~Show~~: $O(bIA) \vee O(bIB)$	DD
5.	~~Show~~: $inA \supset (O(bIA) \vee O(bIB))$	⊃I
6.	inA	Ass
7.	$O(bIA)$	⊃E,2,6
8.	$O(bIA) \vee O(bIB)$	∨I,7
9.	~~Show~~: $inB \supset (O(bIA) \vee O(bIB))$	⊃I
10.	inB	Ass
11.	$O(bIB)$	⊃E,3,10
12.	$O(bIA) \vee O(bIB)$	∨I,11
13.	$O(bIA) \vee O(bIB)$	∨ ⊃,1.5.9

A first diagnosis

- Premises (2)-(4) are all true in the scenario described by Kolodny and MacFarlane:

(2) If the miners are in shaft A, we ought to block shaft A.

(3) If the miners are in shaft B, we ought to block shaft B.

(4) Either the miners are in shaft A or they are in shaft B.

- However, the conclusion of the argument is false:

(5) Either we ought to block shaft A or we ought to block shaft B.

- In the formal reconstruction, the conclusion logically follows from the premises by *modus ponens*, ∨-introduction and proof by cases (Kolodny and MacFarlane call it ∨-elimination).
- So, if the formal reconstruction is correct, then, given that the premises are true and the conclusion false, (at least) one the inference rules must fail to preserve truth, namely
  - either *modus ponens* is not valid,
  - or ∨-introduction is not valid,
  - or proof by cases is not valid.

Rejecting ∨-introduction

- Suppose we reject the validity of ∨-introduction:

$$\begin{array}{l} \varphi \\ \therefore\ \varphi \vee \psi \end{array}$$

- If ∨-introduction is not valid, premises (2)-(4) do not logically imply conclusion (5) in the formal reconstruction:

(2) If the miners are in shaft A, we ought to block shaft A.

(3) If the miners are in shaft B, we ought to block shaft B.

(4) Either the miners are in shaft A or they are in shaft B.

(5) Either we ought to block shaft A or we ought to block shaft B.

A version of the paradox without ∨-introduction

- The problem is that the paradox can be reinstated without using ∨-introduction.
- Indeed, (6) and (7) are both true in the scenario described by Kolodny and MacFarlane:

(6) If we ought to block shaft A, then we ought to block at least one shaft.

(7) If we ought to block shaft B, then we ought to block at least one shaft

- Suppose we add them to our original premises:

(2) If the miners are in shaft A, we ought to block shaft A.

(3) If the miners are in shaft B, we ought to block shaft B.

(4) Either the miners are in shaft A or they are in shaft B.

- Suppose the miners are in shaft A. By (2) and *modus ponens*, it follows that we ought to block shaft A. By (6) and *modus ponens*, it follows that we ought to block at least one shaft.
- Suppose the miners are in shaft B. By (3) and *modus ponens*, it follows that we ought to block shaft B. By (7) and *modus ponens*, it follows that we ought to block at least one shaft.
- Thus, by proof by cases it follows that

(8) We ought to block at least one shaft.

- But (8) is clearly false, since we ought to block neither shaft. Yet, we reached the conclusion without making use of ∨-introduction.

Rejecting proof by cases

- Suppose we reject the validity of proof by cases:

  $\varphi \vee \psi$
  if φ, then ξ
  if ψ, then ξ
  $\therefore \xi$

- Then we block both versions of the paradoxical argument we described, since the both rely on proof by cases.
- Still, according to Kolodny and MacFarlane, this is not sufficient to block yet another version of the paradoxical argument.

A version of the paradox without proof by cases

- Consider this argument:

  (1) We ought to block neither shaft.
  (2) If the miners are in shaft A, we ought to block shaft A.
  (2') Therefore, the miners are not in shaft A.

- We agreed that (1) is true, since it is the safest course in Kolodny and MacFarlane's scenario. Moreover, we agreed that (2) is true, since if the miners are in shaft A, blocking shaft A will save them all. Yet, the conclusion (2') is clearly unwarranted.
- The problem is that from (1)-(2) it seems that we can prove (2') as follows. Suppose that, contrary to (2'), the miners are in shaft A (reductio assumption). Now, if (1) is true, (1') must be true:

  (1') It is not the case that we ought to block shaft A.

  Moreover, from the reductio assumption and (2) it follows by *modus ponens* that we ought to block shaft A, but this contradicts (1'). Therefore, the miners are not in shaft A.
- The proof of the unwarranted conclusion (2') does not rely on proof by cases, so rejecting proof by cases does not help to block this version of the paradox.

Rejecting *modus ponens*

- While only the first paradoxical argument relies on $\vee$-introduction, and only the second relies on proof by cases, all three paradoxical arguments rely on *modus ponens*.
- Thus, given that the premises are true, *modus ponens* is likely to be responsible for the unwarranted conclusion in each of the three arguments.
- Thus, Kolodny and Macfarlane conclude that we should reject the validity of *modus ponens*.

Why *modus ponens* fails with *ought*

- Kolodny and MacFarlane's considerations suggest that *modus ponens* is the villain in the paradoxical argument with *ought*.
- The clues suggest that the paradoxical conclusion is derived because the application of *modus ponens* in the argument somehow fails to preserve truth, and thus takes us from true premises to a false conclusion.
- Notice that the validity of *modus ponens* for conditionals that do not themselves contain conditionals has never been challenged. McGee (1985) claims that "there is every reason to suppose that, restricted to such conditionals, *modus ponens* is unexceptionable."
- Kolodny and MacFarlane now suggest that *modus ponens* is not unrestrictedly valid for simple conditionals either.
- Clearly, if the charge against *modus ponens* has to stick, it must be explained why the application of *modus ponens* in the paradoxical argument fails to preserve truth.

The plan

- Kolodny and MacFarlane's plan to explain why the application of *modus ponens* fails in the paradoxical argument is this:
  1. provide a semantics for *ought*,
  2. provide a semantics for indicative conditionals,
  3. show how the interaction of these semantics causes *modus ponens* to fail in the paradoxical argument.

The semantic framework

truth at a point of evaluation

- The basic notion in Kolodny and MacFarlane's semantic framework is *truth at a point of evaluation*.
- A *point of evaluation* is a pair (w, i), where w is a possible world-state (representing epistemic possibilities), and i is an information state (a set of possible world-states).
- A *possible world-state*, being an epistemic possibility, may be a world in which water is not H_2O or in which Hesperus is not identical to Phosphorus. (So, the set of possible world-states includes worlds that are not metaphysically possible).
- An *information state* (represented as a set of possible world-states) is the set of alternative possibilities which, given what is known, might be the actual world.

The semantic framework

truth

- Assume a *context c* consists, among other things, of a world and a relevant information state.
- For any context c, let
  - i_c be the information state relevant in c, and
  - w_c be the world of c
- A possible way of defining truth *simpliciter* in this semantic framework is this:
  - a sentence S uttered in a context c is *true* iff S is true at w_c, i_c.
- (This is not Kolodny and MacFarlane's final way of defining truth, since they opt for a relativistic definition, but it will do for our purpose).

Polysemic modals

- Modal expressions, namely words like "necessary", "possible", "must", "may", "ought", etc., are notoriously polysemic.
- For example, sentence (9) is naturally understood as the claim that, in view of what I know, I cannot exclude that Mexico City has more inhabitants than Tokyo (epistemic possibility):

(9) It is possible that Tokyo has more inhabitants than Mexico City.

- Sentence (10) is naturally understood as the claim that the actual physical laws exclude that a body travels at a speed greater than 300000 km per second (physical possibility):

(10) It is not possible that a body travels at a speed greater than 300000 km per second.

- Sentence (11) is naturally understood as the claim that the actual state laws exclude that one is prime minister if one has been sentenced to more than two years (deontic possibility):

(11) It is not possible for one to be prime minister if one has been sentenced to more than two years.

- Sentence (12) is naturally understood as the claim that the laws of metaphysics exclude that water $\neq H_2O$ (metaphysical possibility):

(12) It is not possible that water is not H_2O.

Informational modals

- Kolodny and MacFarlane call *informational modals* those modals whose interpretation is sensitive to the information state.
- Typically, epistemic modals are sensitive to the information state.
- If "possible" in (9) is understood epistemically, the truth of (9) depends on the information state we are assuming:

(9) It is possible that Tokyo has more inhabitants than Mexico City.

- For example, if I do not know the number of inhabitants of Tokyio and Mexico City, (9) is true relative to my information state. If I do, (9) is false relative to my information state.

The selection function f

- The interpretation of informational modals, for Kolodny and MacFarlane, depends on a contextually provided *selection function*.
- Their selection function, not to be confused with Stalnaker's, maps an information state to a set of worlds.
- In the case of epistemic modals, the selection function is the identity function, which maps an information state to itself.
- In the case of deontic modals, the selection function maps an information state to the set of worlds that are as *deontically ideal as possible*, given that information.

An illustration

- For example, in the case of the miners, we do not know which shaft they are in, but we know that if we block one of the shafts we risk killing them all, while if we block neither shaft we kill at most one miner.
- So, the best course of action relative to the information we have is one in which we block neither shaft.
- Thus, the selection function, applied to our information state, will yield a set of worlds in which we block neither shaft.
- Now, suppose we knew that the miners are in shaft A. Then, the best course of action, given this information, would be one in which we block shaft A, since in this way we save them all.
- Thus, the selection function, applied to an information state that includes the info that the miners are in shaft A, will yield a set of worlds in which we block shaft A.

The semantics of *ought*

- The semantics of *ought* may now be stated as follows.
- The modal *ought* is represented by the operator $\Box_d$, where d stands for a contextually provided deontic selection function.
- K&M assume the following truth-conditions for sentences of the form $\Box_d\varphi$:
  - $\ulcorner\Box_d\varphi\urcorner$ is true at the world w relative to the information state i iff for all $w' \in d(i)$, φ is true at w', i.
- (Notice: in the definition, d is ambiguous. On the right-hand side of the biconditional d is the contextually provided selection function and on the left-hand side d is an expression which refers to the contextually provided selection function).

An illustration

- Sentence (13) is represented as (13'):

(13) We ought to block shaft A.

(13') $\Box_d$ we block shaft A

- Formula (13') is true relative to a world w and an information state i iff for all $w' \in d(i)$, "we block shaft A" is true in w'.
- In the context described by Kolodny and MacFarlane, the deontic selection function d, applied to the relevant information state, yields a set of worlds in which we block neither shaft.
- Thus, formula (13') is true relative to the world of this context and the information state of this context iff "we block shaft A" is true in every world in which we block neither shaft.
- Since "we block shaft A" is clearly false in these worlds, it follows that (13) is false in the context described by Kolodny and MacFarlane.
- For similar reasons, sentence (14), and thus sentence (5), is false in the same context:

(14) We ought to block shaft B.

(5) Either we ought to block shaft A or we ought to block shaft B.

Sensitivity to the information state

- Notice that, if we add the information that the miners are in shaft A to the information state described by Kolodny and MacFarlane, the truth value of (13) changes:

(13) We ought to block shaft A.

- If we add the information that the miners are in shaft A to the original information state, the deontic selection function d, applied to the state we obtain, yields a set of worlds in which we block shaft A.
- It follows that sentence (13) is true relative to this new information state, since "we block shaft A" is true in all the deontically ideal worlds relative to this state.

Insensitivity to the information state

- According to the semantics proposed by Kolodny and MacFarlane, deontic modals are sensitive to the information state: their truth-value, as we have just seen, may be different at different information states.
- For simple sentences like (15), on the other hand, this is not the case:

(15) The miners are in shaft A.

- The truth value of sentence (15) at w, i depends only on what is the case at w: (15) is true at w, i iff the miners are in shaft A in w.

Semantics for indicative conditionals

- The semantics of indicative conditionals is this (in a first approximation):
  - $\ulcorner [\mathit{if}\,\varphi]\psi \urcorner$ is true at w, i iff ψ is true at w, i', where $i' = \{w' \in i \mid \varphi \text{ is true at } w', i\}$.
- Thus, an indicative conditional is true at a world w relative to an information state i iff the consequent is true at w relative to the information state which consists of the set of worlds of i in which the antecedent is true.
- (Following Kratzer, Kolodny and MacFarlane assume that the consequents of conditionals are always modalized, even when an explicit modal is not present. In the case of indicative conditionals, they assume that the covert modal operator is an epistemic necessity operator).

An illustration

- Sentence (2) is represented as (2'):

(2) If the miners are in shaft A, we ought to block shaft A.

(2') [if the miners are in shaft A] $\Box_d$ we block shaft A

- (2') is true relative to w, i iff "$\Box_d$ we block shaft A" is true at w, i', where i' is the set of worlds $\in i$ in which the miners are in shaft A iff for all the worlds $w' \in d(i')$ "we block shaft A" is true in w'.
- In the context described by Kolodny and MacFarlane, the information state i_c of the context contains both worlds in which the miners are in shaft A and worlds in which the miners are in shaft B.
- Thus, in order for (2') to be true in the world of the context relative to i_c, "$\Box_d$ we block shaft A" must be true at w, i', where i' is the set of worlds $\in$ i in which the miners are in shaft A.
- By the semantics of the deontic operator $\Box_d$, "$\Box_d$ we block shaft A" is true at w, i' iff "we block shaft A" is true in all the worlds in $d(i')$.
- Since i' only contains worlds in which the miners are in shaft A, the worlds that are deontically ideal as possible relative to i' are worlds in which we block shaft A.
- Clearly, "we block shaft A" is true in these worlds, so (2) is true in the context described by Kolodny and MacFarlane.

Validity in information state semantics

- In order to show why the semantics proposed by Kolodny and MacFarlane invalidates modus ponens, we need to define what "valid" means in a semantics in which truth is relative to a world and an information state.
- A natural way of defining validity for this semantics is this:
  - an argument is *valid* iff there is no information state i and world $w \in i$ such that the premises are all true at w, i and the conclusion is false at w, i.
- The condition that $w \in i$ amounts to requiring that, for an argument to be valid, this condition must be satisfied:
  - there is no information state and no world which, on the basis of that information state, might be the actual world, at which the premises are true and the conclusion false.

Why *modus ponens* fails

- The reason why *modus ponens* fails for indicative conditionals is this:
  - by Kolodny and Macfarlane semantics, indicative $\ulcorner if\, \varphi, \psi \urcorner$ is true at w, i on the condition that ψ is true at w relative to the information state i' obtained by excluding from i the worlds in which φ is false.
  - However, *if the truth value of ψ depends on the information state*, this condition may be satisfied even though φ is true and ψ is false at w, i.
- Let's see why this is indeed what happens in the case of the miners described by Kolodny and Macfarlane.

Illustrating why *modus ponens* fails

- As we just saw, in the case of the miners described by Kolodny and Macfarlane, sentence (2) is true relative to the information state i in which we don't know which shaft the miners are in, since in the deontically ideal worlds relative to $i' = (i+$the miners are in shaft A), we block shaft A.

(2) If the miners are in shaft A, we ought to block shaft A.

- If, unknown to us, the miners are in shaft A, (15) is also true relative to i (remember, the truth of (15) does not depend on the information state):

(15) The miners are in shaft A.

- However, as we saw, sentence (13) is false relative to the information state i, since the deontically ideal worlds relative to i are worlds in which neither shaft is blocked:

(13) We ought to block shaft A.

- Thus *modus ponens* is invalid.

References

- Kolodny N, MacFarlane J. (2010) “Ifs and oughts”, *The Journal of Philosophy*, 107:3, pp. 115-143.