EuJAP | Vol. 13, No. 2, 2017
UDK
1 SWINBURNE, R.
164.031

# DIFFERENT ARGUMENTS, SAME PROBLEMS. MODAL AMBIGUITY AND TRICKY SUBSTITUTIONS*

RAFAL URBANIAK
Ghent University
University of Gdansk

*ABSTRACT*

*I illustrate with three classical examples the mistakes arising from using a modal operator admitting multiple interpretations in the same argument; the flaws arise especially easily if no attention is paid to the range of propositional variables. Premisses taken separately might seem convincing and a substitution for a propositional variable in a modal context might seem legitimate. But there is no single interpretation of the modal operators involved under which all the premisses are plausible and the substitution successful.*

**Keywords:** *Church-Fitch paradox, futura contingentia, modal logic, modal operators, propositional quantification, Swinburne's modal argument*

## 1. Introduction

Certain arguments use modalities in close, but different meanings. This might lead to the situation in which premisses taken separately seem convincing (and substitution for a propositional variable might seem legitimate), but nevertheless, no single interpretation of the modal operators involved makes all the premisses plausible and the substitution legitimate.

While it's difficult to *a priori* point to a wider class of arguments in which the problem arises, the issue might be more common than it might seem: at least in philosophical arguments it occurs in quite different contexts. This suggests that philosophers should be on the lookout for this type of error whenever a philosophical argument involving both modalities and propositional quantification is involved.

*Received: 09.04.2016.
Accepted: 14.12.2017.

The goal of this paper is to diagnose this problem in a few fairly well-known philosophical arguments, which normally aren't discussed together, and whose similarity hasn't been previously noticed:

1. Swinburne's modal argument for the existence of the soul;
2. a logical argument for fatalism;
3. the Church-Fitch paradox.

The goal of discussing arguments concerning quite different topics is to emphasize that the flaw isn't too topic-dependent. The arguments were chosen because they are well-known, they concern different topics and yet they all commit the same fallacy.

From the assumption that it is logically possible that a human being survives the destruction of their body and a few additional modal premisses Richard Swinburne infers the actual existence of souls. Various variants of the argument against future contingents rely on modal and temporal premisses and seem to lead to the conclusion that there are no future contingent events. The Church-Fitch paradox leads to the conclusion that the existence of unknown truths excludes all truths being knowable. I start with presenting the first two arguments, then I argue that whatever appearance of soundness they have, they owe it to the ambiguity of the modal operators involved and lack of attention to propositional quantifier range. Then I describe the third argument and point out a similar issue with it.

While the arguments for the sake of clarity and brevity are to some extent formalized, the main point is not about the formal tools, but rather about their misuse in representation of the underpinning philosophical intuitions.

## 2. Swinburne's modal argument

Let's start with the original formulation of the argument. The argument in its fullest version can be found in (Swinburne 1986, ch. 8). It also occurs in (Swinburne and Shoemaker 1984, ch. 2). In (Swinburne 1996) the author develops a defense of the modal argument against certain objections raised in the literature of the subject.

First, some abbreviations. $\Diamond$ is the possibility operator, $\Box$ is the necessity operator, $\wedge$ is the conjunction symbol, $\rightarrow$ is material implication, $\Rightarrow$ is the logical/definitional implication, $\leftrightarrow$ is material equivalence, $\Leftrightarrow$ is the logical/definitional equivalence, and $\neg$ is the negation symbol.

The key difference between the single arrow symbols and double arrow symbols is that the former are connectives in the object language, while the latter are meta-linguistic. Moreover, material equivalence only says that it is not the case that one side is true and the other false, while the definitional/logical equivalence requires that it is necessarily so, and allows for the substitution of equivalents in modal contexts.

I customized propositional constants for mnemonic purposes.

> C ⇔ Swinburne is a **C**onscious person and exists in 1984.
> D ⇔ Swinburne's body is completely **D**estroyed in the last instant of 1984.
> S ⇔ Swinburne has a **S**oul in 1984.
> E ⇔ Swinburne **E**xists in 1985.

Swinburne introduces a variable $p$ that is supposed to range over propositions of a specific sort: "p ranges over all consistent propositions compatible with $C \wedge D$ and describing 1984 states of affairs." (Swinburne 1996, 69) We'll work with **T** as the underlying modal logic (that is, apart from distributing $\Box$ over implication, we have reflexivity ($\Box p \rightarrow p$) which requires that whatever is necessary is true.

The first premiss of the argument is contingent. It says that Swinburne is a conscious person and exists in 1984:

(1) $C$

The second premiss states that for any sentence about 1984 compatible with C and D it is possible that Swinburne survives the destruction of his body, and yet that his compatible sentence is true:

(2) For all $p$, $\Diamond(C \wedge D \wedge p \wedge E)$

The third premiss says that it is not possible (at least for Swinburne) to survive the complete destruction of his body if he doesn't have a soul (an immaterial part):

(3) $\neg\Diamond(C \wedge D \wedge \neg S \wedge E)$

Premiss (2) says that any sentence compatible with $C \wedge D$ and describing 1984 states of affairs is compatible with $C \wedge D \wedge E$ but premiss (3) says that $\neg S$ is not compatible with $C \wedge D \wedge E$. Therefore, $\neg S$ is not a sentence that is compatible with $C \wedge D$ and describes 1984 states of affairs. Or, in other words, premisses (2) and (3) together entail that $\neg S$ is not within the range of $p$. But if $\neg S$ is not compatible with $C \wedge D$, then $C \wedge D$ entails S. But D doesn't have any impact on the truth of S, and so, if $C \wedge D$ entails S, then so does C alone.

The argument has been developed into a fully formalised form and reformulated into a version immune to what was considered the main objection put forward in (Zimmerman 1991; Alston and Smythe 1994; Stump and Kretzmann 1996). Full details of the construction and a longer discussion of known objections can be found in (Urbaniak and Rostalska, 2010). Here I just present the final effect.

To proceed with our analysis, we need three more abbreviations.

$84(p) \Leftrightarrow$ it is a fact about 1984, that $p$

By $84(p)$ we only mean that $p$ states something about an event or a state of affairs in 1984 and it does not state anything about an event or state of affairs "outside of" 1984. What is also important, a sentence does not have to be true in order to be about 1984. The notion of being about 1984 is a bit vague, but in fact we do have decent intuitions about whether a sentence is (purely) about 1984. So, for instance, we exclude sentences like:

It is the case in 1984 that in 1985 Swinburne will not exist.

Although, in a way, this sentence is about 1984, it is not purely about 1984, because it clearly implies a contingent sentence about 1985. However, both sentences:

Swinburne is purely material in 1984.
Swinburne is not purely material in 1984.

seem, on the face of it, to be purely about 1984. While *prima facie* it might seem that 84(-) is a predicate, it is intended as a connective: "*it is a claim about 1984 that ...*" is supposed to be completed by a sentence, not a name thereof. Sometimes, in informal discussion I will simplify the discourse by speaking as if I was talking about a predicate, but in such cases nothing in the discussion prevents reformulation in which it is made explicit that 84(-) is a connective.

Now, we also add a piece of notation for expressing the property of *being true about 1984*:

(4) $\mathrm{tr}84(p) \Leftrightarrow 84(p) \wedge p$

The third abbreviation is:

(5) $\Diamond(p \wedge \mathrm{C} \wedge \mathrm{D}) \Leftrightarrow p$ is compatible with $\mathrm{C} \wedge \mathrm{D}$.

It may seem slightly unclear what sort of compatibility Swinburne has in mind. He emphasises that it is the same notion as that of logical coherence, quite explicitly denying that there is a separate "metaphysical" kind of necessity: "…the contrast is misleading. For not merely is the necessity of both kinds equally hard, but has the same nature - the necessary is that which holds in all possible worlds, where 'possible' means 'coherently describable' " (Swinburne 1986, 314).

Now, we can carefully state the evolved version of the argument. The first premiss only states that Swinburne is alive and conscious in 1984:

(6) C

The second premiss is a modification of Swinburne's original second premiss. It also captures the assumption about the range of variables that was mentioned but not included in the formula. At the first stab we might want to formalize the assumption as follows:

(7) For all $p$, $[84(p) \wedge \Diamond(p \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge p \wedge E)]$

This, however, would leave the argument unnecessarily open to the following objection. Eliminate the universal quantifier and substitute M for $p$, where

M⇔Swinburne is purely material in 1984 for $p$.

84(M) seems intuitively true: Swinburne's being purely material in 1984 is a fact about 1984. Moreover, $\Diamond(M \wedge C \wedge D)$ also seems true, unless we want to decide the issue at question beforehand: it is at least possible that Swinburne is purely material and conscious in 1984, and his body is destroyed in the last instant of 1984. This would make the antecedent of the resulting substitution at least strongly plausible. The consequent — $\Diamond(C \wedge D \wedge M \wedge E)$ — however, would say that it is possible that Swinburne is conscious and purely material in 1984, his body is destroyed in the last instant in 1984, and yet he manages to survive into 1985. This doesn't seem plausible, and so (7) could be argued to entail a substitution which isn't very convincing.

Now, using tr84 instead of 84 we obtain:

(8) For all $p$, $[\mathrm{tr}84(p) \wedge \Diamond(p \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge p \wedge E)]$

(8) says about *any* proposition $p$ that if it is true, purely about 1984, and compatible with $C \wedge D$, it is compatible not only with the claim that Swinburne is conscious and alive in 1984 and his body is destroyed in the last moment of 1984, but also compatible with the claim that Swinburne is conscious and alive in 1984, his body is destroyed in the last moment of 1984 and yet he survives and exists in 1985.

(8) is not susceptible to an objection analogous to the one that we just put forward against (7). For say we eliminate the universal quantifier and substitute M for $p$. To argue that this substitution is false, we need to argue that its antecedent is true. But the first conjunct now reads tr84(M) — $84(M) \wedge M$ — and while in the previous argument we only needed the assumption that being material and conscious is at least possible, now we would need the assumption that Swinburne indeed was purely material in 1984. But insisting that this is the case already decides the issue to be decided by the argument. To undermine a premiss, we'd be arguing that it is false, because the conclusion of the whole argument is — not an interesting criticism at all.

Premiss three is exactly the same as in the original argument:

(9) $\neg\Diamond(C \wedge D \wedge \neg S \wedge E)$

It is obviously equivalent to:

(10) $\Box(C \wedge D \wedge E \rightarrow S)$

(10) says that necessarily, if Swinburne is conscious and alive in 1984, his body is destroyed in the last moment of 1984 and yet he exists in 1985, he has a soul in 1984. It is meant to capture the intuition that to survive the destruction of one's body, one has to have a soul.

The next premiss says that 'Swinburne does not have a soul in 1984' is purely about 1984:

(11) $84(\neg S)$

and another one claims that if C and D necessarily entail S, then so does C:

(12) $\Box(C \wedge D \rightarrow S) \rightarrow \Box(C \rightarrow S)$

That is, if the facts that Swinburne exists and is conscious in 1984 and that his body is destroyed in the last moment of 1984 entail that Swinburne has a soul in 1984, the fact that Swinburne's body is destroyed in the last moment of 1984 has no relevance for this conclusion and the very fact that Swinburne is alive and conscious in 1984 already necessitates the fact that Swinburne has a soul in 1984.

These assumptions logically entail S. First, eliminate the universal quantifier from (8), substituting $\neg S$ for p:

(13) $tr84(\neg S) \wedge \Diamond(\neg S \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge \neg S \wedge E)$

from (9) and (13) we obtain:

(14) $\neg[tr84(\neg S) \wedge \Diamond(\neg S \wedge C \wedge D)]$

we apply De Morgan's law to it:

(15) $\neg tr84(\neg S)$ or $\neg\Diamond(\neg S \wedge C \wedge D)$

At this point we split the disjunction into a proof by cases. Suppose $\neg tr84(\neg S)$. By (4) this means that either $\neg 84(\neg S)$ or $\neg\neg S$. But (11) says that $84(\neg S)$. So $\neg\neg S$ and hence S. Suppose on the other hand that $\neg\Diamond(\neg S \wedge C \wedge D)$. In this case we get:

(16) $\neg\Diamond(\neg S \wedge C \wedge D)$

Quite easily we now obtain:

(17) $\Box\neg(\neg S \wedge C \wedge D)$

Since:

(18) $\neg(\neg S \wedge C \wedge D) \Leftrightarrow [(C \wedge D) \rightarrow S]$

and because we can substitute logically equivalent expressions within the scope of modal operators, we can infer:

(19) $\Box((C \wedge D) \rightarrow S)$

Now, we apply *modus ponens* to (12) and (19) and obtain:

(20) $\Box(C \rightarrow S)$

With (6), by reflexivity for $\Box$ and *modus ponens*, this entails:

(21) S

Either way we obtain S, which completes the argument.

## 3. Assessing Swinburne's argument

The argument might be criticised for being epistemically circular in the following sense. Basically, (8) says that no *true* proposition compatible with Swinburne's being alive and conscious in 1984 and his body being destroyed excludes the possibility of him surviving the destruction of his body. This is a fairly strong claim, because it is equivalent to the claim that *any* proposition about 1984 compatible with C∧D, which excludes the possibility of Swinburne's survival (while C∧D), is already false.

To see the equivalence, unpack the expression in the scope of the quantifier in (8) as follows:

$$84(p) \wedge \mathrm{p} \wedge \Diamond(p \wedge C \wedge D) \rightarrow \Diamond(C \wedge D \wedge p \wedge E)$$

It is now a matter of purely propositional manipulation (contraposition, really) to see that this is equivalent to:

$$84(p) \wedge \Diamond(p \wedge C \wedge D) \wedge \neg \Diamond(C \wedge D \wedge p \wedge E) \rightarrow \neg p$$

Thus, for instance, if one believes that M (⇔ Swinburne's is purely material in 1984) excludes such a possibility, and is compatible with C∧D, one is committed also to the falsity of M. But if this is the case, by accepting (8) we already seem to have a firm philosophical position on the issue.

A way out seems to be to say that no sentence purely about 1984 is incompatible with Swinburne's survival in 1985 because sentences purely about 1984 don't entail anything about 1985. So, no such sentence, even if true, could exclude Swinburne's survival of the destruction of his body. So to avoid the difficulty from the previous paragraph, one needs to interpret compatibility using a Humean notion of necessity on which no truth about one time necessitates anything about some other time, presumably Swinburne's notion of logical necessity.

Alas, the problem is that if we assume that our notion of compatibility is purely logical, we now have a reason to reject premiss (9) which says that

Swinburne cannot survive the complete destruction of his body and continue to exist in 1985 if he doesn't have a soul in 1984. After all, as Humean logical modality is involved, we have no reason to think that a sentence purely about 1984 should be analytically incompatible with any sentence about 1985. In other words, to accept (9) we have to use a stronger, presumably metaphysical interpretation of modality on which not having a soul in 1984 is incompatible with surviving the destruction of the body at the end of 1984.

Swinburne argued (in personal communication) that the truth of (8) is available even to children when he tries to explain the argument to them. They tend to agree when faced with statements like 'Look, it is at least logically possible, whatever else is true and doesn't exclude us being conscious and our bodies being completely destroyed, that we survive the complete destruction of our bodies'. There are two points to make to explain these intuitions away. First, there is a notion of possibility on which (8) comes out true, but which falsifies (9). It is quite likely that some people, when faced with the sentence quoted in this passage, use this notion to assess its truth. Then, when they're faced with the informal reading of (9), they use quite a different metaphysical notion of possibility not noticing the difference. Second, there is an important scope distinction to be kept in mind. On one reading, the quoted sentence says exactly what (8) says, and yields a rather strong statement. On the other reading, it rather says that no matter what is true about 1984 and doesn't exclude C and D, it is still possible to survive the complete destruction of one's body. On the second reading, anyone who admits $\Diamond(C \wedge D \wedge E)$, a rather weak claim of mere logical possibility is committed to this claim. On this reading, however, the claim is too weak to constitute a premiss of a valid modal argument.

The issue can be rephrased in terms of substitutions: on one reading, substituting $\neg S$ for *p* when eliminating the quantifier yields a false antecedent, but we have no reason to accept the premiss, and on another, we have a reason to accept both the premiss and its substitution, but we have no reason to think that the consequent of the resulting sentence is false.

## 4. Arguing against future contingents

Aristotle in the IXth chapter of *On Interpretation* considers an argument for the non-existence of propositions about future contingent events. The argument is supposed to defend *fatalism*, the view that each future event will take place necessarily. The view that one can prove fatalism on merely logical grounds is called *logical fatalism*. I put well-known interpretative issues related to Aristotle's formulation aside (see however Øhrstrøm and Hasle 1995, 6-109) and look at a streamlined version of one of the best formulations of an argument for logical fatalism as developed by Prior and Rescher (see Prior 1967, 119-121).

- Formula F$yp$ is read *in y units of time it will be the case that p*.
- Analogously we read the formula P$yp$ as *y units of time ago it was the case that p*.
- Formula T$ap$ reads *it is true at time a that p.*

Before we move to the argument itself, some preliminary inference rules are needed:

(RT) If $\vdash p$, *then* $\vdash \mathrm{T}ap$, for all $a$

(TC) $\mathrm{T}a(p \rightarrow q) \vdash \mathrm{T}ap \rightarrow \mathrm{Tab}$

(RD) If $\vdash p$, then $\vdash \Box ap$, for all $a$

(DC) $\Box a(p \rightarrow q) \vdash \Box ap \rightarrow \Box aq$

Roughly, (RT) says that whatever is provable is true at any time $a$, and (TC) says that truth at any time $a$ distributes over implication. (RD) says that whatever is provable is necessary at any time $a$, and (DC) states that being necessary at $a$ distributes over implication. All of these principles seem quite convincing.

We start the argument with a premiss saying that $p$ will be true in $n$ units of time just in case it was $m$ units of time ago true that $p$ would be true in $m+n$ units of time, and another premiss saying that if it is true at $a$ that $p$ was true $m$ units of time ago, it is necessary that $p$ was true $m$ time units ago:

(P1) $\mathrm{F}np \leftrightarrow \mathrm{P}m\mathrm{F}(m+n)p$

(P2) $\mathrm{T}a\mathrm{P}mp \rightarrow \Box a\mathrm{P}mp$

(P2) is supposed to capture the intuition that the past is necessary and cannot be changed. Now, (RT) and (TC) allow us to first add T$a$ in front of the left-to-right implication from (P1) and then distribute it over the implication:

(AP2) $\mathrm{T}a\mathrm{F}np \rightarrow \mathrm{T}a\mathrm{P}m\mathrm{F}(m+n)p$

Let's now substitute F$(m+n)p$ for $p$ in (P2):

(AP3) $\mathrm{T}a\mathrm{P}m\mathrm{F}(m+n)p \rightarrow \Box a\mathrm{P}m\mathrm{F}(m+n)p$

(AP2) with (AP3) give us:

(AP4) $\mathrm{T}a\mathrm{F}np \rightarrow \Box a\mathrm{P}m\mathrm{F}(m+n)p$

Now we take the right-to-left implication from (P1) and use (RD) and (DC) to introduce and distribute the necessity operator:

(AP5) $\Box a\mathrm{P}m\mathrm{F}(m+n)p \rightarrow \Box a\mathrm{F}np$

Finally, we put (AP4) and (AP5) together, obtaining the conclusion of the argument:

(AP6) $\mathrm{T}a\mathrm{F}np \rightarrow \Box a\mathrm{F}np$

This conclusion says that whenever it is at *a* true that *p* will take place in *n* time units, it is necessarily the case.

Having described the arguments, let's focus on identifying the type of error that makes it seem plausible.

## 5. Assessing argument about future contingents

(RT), (RD) and (P1) seem rather innocent. One could be worried that (TC) and (DC) correspond to axiom K of the standard modal logic, which results in issues related to logical omniscience and some deontic paradoxes. These issues, however, don't seem to carry over to the temporal reading. As for the former, while perhaps it is a problem that actual agents' knowledge is not closed under logical consequence, there is no good argument to the effect that being true at a time shouldn't be so closed. As for the latter, most notable paradoxes are the Good Samaritan paradox and Ross's paradox. The Good Samaritan arises when one formalizes *It ought to be the case that Jones helps Smith who has been robbed* as $O(h\wedge r)$ and then uses K to infer $Or$, that Smith should be robbed. Part of the issue is that the formalization does violence to the original premiss, which, come to think about it, doesn't say that it ought to be the case that Jones helps Smith and that Smith has been robbed. But even if we're worried about conjunction elimination in deontic contexts, it doesn't seem to be problematic when applied to being true at a time.

Ross's paradox, on the other hand, arises when K is used to deduce $O(m\vee b)$, that *it ought to be the case that the letter is mailed or burned,* from $Om$, the claim that *it ought to be the case that the letter is mailed.* Again, while there might be reasons to think that obligation is not closed under disjunction, there is no analogous intuition about being true at a time.[1]

[1] I am grateful to an anonymous referee for pressing me on the issues with these prima facie suspicions about (TC) and (DC).

So it seems the most troubling move is the one captured by (P2). How exactly should we understand the modality in (P2)?

Alas, on this reading, no one who doesn't already accept causal determinism will accept the premiss itself. For then (P2) says that whatever happened in the past was causally determined, which is not something that someone who believes certain past human choices were made freely would accept.

So perhaps we should focus on a somewhat different reading, according to which it is not *being causally determined*, but rather *there being at time a no way of influencing the truth-value of p*. On this reading, the premiss only seems to say that whatever is in the past can no longer be changed. And indeed, we do seem to have this intuition if *p* is a simple sentence clearly about the past. For instance, if it so happened that student X failed his logic exam, nothing in the future can be done to change this fact from the past.[2]

But now, the key move is from (P2) to (AP3), where for *p* we substituted F($m$+$n$)$p$. As pointed out already by Ockham (see Boehner and Ockham 1945), claims such as *it was the case m units of time ago that in m+n units of time p will be the case* are not strictly about the past. While an indeterminist might still have the intuition that we can do nothing to change past events (even though some of the past choices were free at the time when they were made), she will definitely deny that such spuriously past-tensed descriptions really describe a past event.

The problem is clear: as far as in this interpretation (P2) is convincing for sentences not about the future, if we allow substitutions of sentences involving the future it turns out to build in the intended conclusion: the necessity of all future events. (Having said that, developing a formal semantics for temporal logic which captures the distinction between legitimate and illegitimate substitutions is quite a task - see for example the discussion in Prior 1967.)

This can be avoided by restricting the range of legitimate substitutions (we did the same thing when making the range of quantifiers explicit in Swinburne's argument). Say by D<$ap$ we mean *the truth value of p does not depend on any facts occurring after a*, (P2) should be replaced with

(P2′) $\mathrm{D}{<}ap \rightarrow (\mathrm{T}a\mathrm{P}mp \rightarrow \Box a\mathrm{P}mp)$

Switching to (P2′) avoids the objection we just raised: for (P2′) now says that if *the truth-value of p doesn't depend on any facts occurring after a,* then if it is true at *a* that *p* was true *m* units of time ago, it is also necessary at *a* that *p* was true *m* units of time ago.

---

[2] Note: the argument assumes the past cannot be changed. Already this is a non-trivial philosophical assumption excluding, for instance, time-travel. But let's play along. Also, there might be logical reasons to to take the possibility of time-travel too seriously (see Urbaniak 2007).

The problem now is that the conclusion of the whole argument with (P2′) in place of (P2) no longer follows because it clearly is not the case that $\mathrm{D}{<}a\mathrm{F}(m+n)p$.

Why did we buy into (P2) in its full strength to start with, then? For one thing, plain carelessness and not thinking about whether accepting (P2) with all its tricky substitutions captures our intuition about the *past* being irreversible. But I think there's an associated reason, that has to do with further confusion with another modality. We are so used to thinking about alethic modalities such as logical or metaphysical necessity and to the ultimate validity of logical principles involving such modalities, that we are without much hesitation willing to accept all substitutions of what seems to us to be a valid general principle about a modality. This however should not be the case with temporal modalities, which turn out to be a bit more tricky.

So we observed a situation where in a modal context our intuitions about which substitutions should be admitted lead us astray. The situation, it seems, isn't very unique. Let's take a look at yet another case.

## 6. Church-Fitch paradox

Here's another well-known argument that turns around a tricky substitution mixed with a modal context. The goal of the argument is to show that one cannot consistently claim that all truths are knowable and that there is at least one unknown truth. Let's start with formulating the premisses, next we'll see how the conclusion follows, and then we'll move to the assessment. Aside from classical logic in the background, the premisses and principles needed for the argument are as follows (K reads *it is known that/one knows that*):

(IK) For some $p$ $(p \wedge \neg \mathrm{K}p)$

(Know) For all $p$, $\vdash (p \rightarrow \Diamond \mathrm{K}p)$

(Distr) $\mathrm{K}(p \wedge q) \vdash \mathrm{K}p \wedge \mathrm{K}q$

(Fact) $\mathrm{K}p \vdash p$

(Contra) If $p \vdash \bot$, then $\Diamond p \vdash \bot$

(IK) says that our knowledge is incomplete, that there is at least one unknown true proposition. (Know) states the knowability of any true proposition. (Distr) allows to distribute knowledge over a conjunction, and (Fact) states the factivity of knowledge: that whatever is known is true. (Contra) says that the possibility claim of a contradictory proposition is already contradictory.

Now the reasoning. (IK) says that there is a true sentence, say $n$, that is not known. (in a formal proof this move would correspond to the elimina-

-tion of the existential quantifier, substituting a fresh propositional constant $n$ for $p$),

(IK2) $n \wedge \neg Kn$

As (Know) contains a universal quantifier, we can eliminate it, substituting any formula whatsoever for $p$. In particular, let's substitute the content of (IK2):

(Know2) $n \wedge \neg Kn \rightarrow \Diamond K(n \wedge \neg Kn)$

Now we can simply apply detachment to (IK2) and (Know2) getting:

(IK3) $\Diamond K(n \wedge \neg Kn)$

Now, we will show that what's in the scope of $\Diamond$ in (IK3) entails a contradiction, which thanks to (Contra) will allow us to deduce contradiction from (IK3) itself. We apply (Distr) to $K(n \wedge \neg Kn)$ and eliminate the conjunction:

(IK4) $Kn$

(IK5) $K\neg Kn$

Now, we apply (Fact) to (IK5) to obtain

(IK6) $\neg Kn$

which together with (IK4) yields a contradiction. This means that by (Contra), a contradiction follows already from (IK3). We obtained the conclusion that a contradiction is possible, whose negation can be proven in a very rudimentary modal logic K. So it seems that (IK) and (Know) exclude each other!

Perhaps, we're to blame (Distr) or (Fact) for the pickle? Well, (Fact) seems like a principle capturing the factivity of knowledge, and isn't independently known to lead to undesired consequences. One might, however, have an issue with (Distr) and axiom K, pointing out that they build logical omniscience into the system (well, Distr does this only partially).

But what would an attempt to solve the paradox by insisting that (Distr) doesn't apply here look like? One would have to deny the inference from (IK3) to (IK4) — that is, one would have to deny that if a subject knows that ($n$ and it is not known that $n$), then it follows that the subject knows that $n$ and they know that it is not known that $n$. This doesn't seem too convincing and would require an independent motivation. Moreover, such a solution would, so to speak, come too late: already (IK3) seems rather absurd, and so something problematic must've preceded the application of (Distr).[3]

[3] Thanks to an anonymous referee for pressing me on the issue of logical omniscience.

It seems that the issues with logical omniscience and those brought up by the Church-Fitch paradox aren't too related — after all, the reasoning involved is very simple and doesn't involve any massive complication that could prevent the subject from grasping the fact, say, that conjuncts follow from a conjunction and so their knowledge is closed at least under eliminating conjunction once.

## 7. Assessing Church-Fitch paradox

Now, the whole trick is made possible by substituting $n\wedge\neg\mathrm{K}n$ for $p$ in the knowability principle (Know), leading to (Know2). And come to think about it, if $n$ is a true but unknown sentence, there is nothing amazing about $n\wedge\neg\mathrm{K}n$ not being knowable. After all, if you know this conjunction, you know $n$, so you at the same time falsify the second conjunct! The whole argument is just a way of explicating this fact.

But of course, you might think that this is a rather cheap shot, because this is not the sort of substitutions you had in mind. You can still accept the spirit of the knowability principle and say: *when I said all truths are knowable I meant actual truths about the non-epistemic world, not some tricky sentences involving claims about what is known, I don't know which tricky sentences involving epistemic operators are and which aren't knowable!* And once this restriction on substitution is made, the argument doesn't fly to far.

Having said this, the challenge of developing a formal framework explaining which substitutions exactly lead to problems and which don't is quite daunting. This is especially so if we want to be less conservative than we are by excluding all formulas involving epistemic operators. The problem generated quite a lot of literature (see for instance Kvanvig 2006 and Salerno 2009). People who think this is an important problem have spent considerable amount of time thinking about this without reaching agreement.

One notable proposal, for instance, is due to Tennant (2002). Take the example of "No thinkers exist". The proposition is consistent, but can't be known to be true (assuming only thinkers can know things). Such sentences (whose knowledge claims are not possible) Tennant calls *anti-Cartesian*. Sentences which are not anti-Cartesian are said to be *Cartesian.* A sentence of the type $Kp$ can be impossible for various reasons: it might be that $p$ itself is inconsistent, it might be the case that judging that $p$ requires the falsity of some consequence of $p$, or it might be that $Kp$ is impossible due do the logical structure of $p$ itself, despite $p$ being consistent. Tennant's antidote to the Church-Fitch paradox is: restrict the knowability claim to Cartesian propositions only.

The approach definitely blocks the paradox, and is definitely common-sensical in spirit: don't use tricky unintended substitutions which due to logical complexities involve you in impossibilities. It also nicely illustrates how rather unexciting the philosophical lessons from taking such paradoxes too seriously can get. At its core, the proposal simply is: *when I say that all truths are knowable, of course I don't mean truths which a priori can't be known for various rather trivial reasons that have to do pretty much with self-reference.* In all fairnes, it's hard to draw any deeper lesson here.

Again, the phenomenon seems to be that initially a formula seems like an adequate formalisation of our pre-formal claim, and only after further formal development which uses tricky unintended and unexpected substitutions it is revealed that the formula in its whole unintended generality wasn't so convincing to start with.

## 8. So what do they have in common?

I surveyed three different arguments about quite different issues, which turned out to share the following features:

(A) **They all employ somewhat unexpected substitution for a propositional variable or for a schematic letter.**

- In the case of Swinburne's argument, premise (8), which roughly captures the intuition that nothing consistent with Swinburne's being conscious prior to the destruction of his body excludes his survival, is used by substituting a claim on the very matter at hand "Swinburne doesn't have a soul" for the propositional variable, and the interplay of various modalities involved in the argument with our shaky intuitions are used to make the argument seem plausible.
- In the case of the logical argument against future contingents, a claim involving tricky reference through time, $F(m+n)p$, is substituted for p in (P2).
- In the Church-Fitch paradox a clearly a priori unknowable claim is substituted for a variable in the knowability principle, which doesn't seem to be intended to be applicable to such claims.

(B) **The substitution is applied to a *prima facie* plausible premiss.**

- In Swinburne's argument the intuition supporting (2) is that at least we should treat the possibility of Swinburne's survival as an open possibility.
- In the fatalistic argument, (P2) at least prima facie is supported by the intuition that past cannot be changed.

- In the Church-Fitch paradox the assumption is an expression of cognitive optimism to the effect that at least in principle any aspect of the world can be the subject of knowledge.

(C) **The substitution is made in a modal context.**

- In Swinburne's argument, as discussed, how the modalities are understood bears on how plausible particular premisses are.
- In the fatalistic argument the modalities involved are temporal, and various considerations arise from different interpretations of what it means for the past to be unchangeable.
- In the Church-Fitch paradox, the modal context is epistemic.

Given the discussion so far, the following claims also clearly hold:

(D) **The arguments lead to very strong conclusions from prima facie innocent premisses.**

(E) **The unclarity about which modality is involved and which substitutions are intended as admissible, given the choice of modality, makes the premisses prima facie plausible and the argument at least initially plausible.**

Problems with these arguments are not at the formal level: they arise at the level of formalisation. The goal of the paper wasn't to show that there are *logical* flaws in the arguments. Quite the opposite: with all assumptions in place, the (semi-)formal arguments are logically correct in the sense that the conclusion follows from the premisses by means of the rules stated. The goal, however, was to show that logically correct arguments can be used to philosophically suboptimal effects, as long as not enough attention is paid to formalization, to the underlying intuitions, and philosophical justifications of the premisses.

Given that the problems are with philosophical moves and not with formal systems, it is difficult to formulate simple general principles that would help to avoid such infelicities. I hope my critical survey, however, will raise the reader’s sensitivity to the meaning of modalities and to the range of propositional variables used in various arguments. These issues, while they might sound quite distant from the point of view of a serious philosopher who wants to argue for a strong philosophical position, cannot be ignored. The devil is in the detail.

**Acknowledgments**

I would like to express my gratitude to two anonymous referees of the journal, whose in-depth comments were extremely useful.

**REFERENCES**

Alston, W. P. and Smythe, T. W. 1994. Swinburne's argument for dualism. *Faith and Philosophy* 11(1): 127–133.

Boehner, P. and Ockham, G. 1945. *The Tractatus de Praedestinatione Et de Praescientia Dei Et de Futuris Contingentibus of William Ockham: With a Study on the Mediaeval Problem of a Three-valued Logic*. New York: Franciscan Inst. St. Bonaventure College.

Kvanvig, J. L. 2006. *The Knowability Paradox*, Oxford: Oxford University Press.

Øhrstrøm, P. and Hasle, P. F. 1995. *Temporal logic: from ancient ideas to artificial intelligence*. Dordrecht: Springer.

Prior, A. N. 1967. *Past, Present and Future*. Oxford: Clarendon Press.

Salerno, J. eds. 2009. *New Essays on the Knowability Paradox*. Oxford: Oxford University Press.

Stump, E. and Kretzmann, N. 1996. An objection to Swinburne's argument for dualism. *Faith and Philosophy*,13: 405–412.

Swinburne, R. 1986. *The Evolution of the Soul*. Oxford: Clarendon Press. Page references to the second edition published in 1997 by Clarendon Press.

_______ 1996. Dualism intact. *Faith and Philosophy* 13(1): 68–77.

Swinburne, R. and Shoemaker, S. 1984. *Personal Identity.* Oxford: Basil Blackwell.

Tennant, N. 2002. *The Taming of the True*, Oxford University Press.

Urbaniak, R. 2007. Time Travel and Conditional Logics. In *Logica 2007 Yearbook*, ed. M. Peli, 247–255.

Urbaniak, R. and A. Rostalska. Swinburne's modal argument for the existence of the soul. *Philo*, 12:73–98, 2010.

Zimmerman, D. W. 1991. Two cartesian arguments for the simplicity of the soul. *American Philosophical Quarterly* 3: 217–226.