MODAL EPISTEMOLOGY

23 September 2004

Erik Weber & Tim De Mey (eds.)
MODAL EPISTEMOLOGY

CONTENTS

Erik Weber & Tim De Mey
Introduction ........................................... 7

Gerald J. Massey
From Conceivability to Possibility: Method or Myth? .......... 9

Filip Buekens
From Conceivability to Possibility: the Normative Account .. 23

Daniel Cohnitz
Why Consistentism Won't Work ........................ 33

Tim De Mey
Mere Logical Possibilities: a Bridge too Far .................. 53

Sören Häggqvist
Modality and Similarity .................................. 63

Gunmar Björnsson
A Naturalist's Approach to Modal Intuitions .................. 77

Ondřej Bueno & Scott Shaikiowski
Modal Realism and Modal Epistemology: a Huge Gap .......... 93
REFERENCES


WHY CONSISTENTISM WON'T WORK

Daniel Cohnitz

Philosophy Department, University of Düsseldorf

Abstract. Consistentism is the doctrine that what is knowable to be (logically, conceptually, metaphysically, physically) possible is nothing but what is not provable to be incompatible with what is knowable to be (logically, conceptually, metaphysically, physically) necessary, or in other words, that ideal negative conceivability entails possibility. In this paper, I will first show that consistentism is at the heart of various modal epistemologies, as well as some modal ontologies. The most famous candidates are David Chalmers' notion of ideal negative conceivability, Bob Hale's necessity based account of modal epistemology, and Peter Menzies' response-dependence analysis of modality. Consistentism is flawed. It is not flawed in the sense that there are some far-fetched counterexamples. It is flawed in the strong sense that it is logically infeasible, and the aim of the paper is to prove this.

1. WHAT CONCEIVABILITY IS

Since Thomas Reid there is an entertaining intellectual game for philosophers to engage in during boring winter evenings: disambiguate the sentence 'It is conceivable that p' in so many ways that eventually one version will not obviously fail to imply 'It is possible that p'. This sport was invented to render Hume's Maxim

That whatever the mind clearly conceives, includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible. (Hume 1768: 32)

1 I would like to thank Stefan Bagusche, Manuel Bremer, Axel Bühlert, Ross Cameron, David Chalmers, Bob Hale, Tim De Mey, Sven Rosenkranz, Marcus Rossberg, Gerhard Schurz, Markus Werning, the audience of my talk at GAP5, the members of Sven Rosenkranz's seminar on modal epistemology at the FU Berlin, and an anonymous referee for helpful comments on earlier presentations of the thoughts presented in this paper. I would also like to thank the Arché Centre for the Philosophy of Logic, Language, Metaphysics and Mind (in particular Fraser MacBride and Avi Heffeman) at St. Andrews for the hospitality during summer 2003 and the opportunity to work in an extraordinary inspiring environment.

2 Reid 1786: 360-379. Already Descartes and Arnaud tried to explicite a philosophical notion of conceivability, but Reid seems to be the first who thought about it systematically.
intelligible. The botany of conceivability has made considerable progress thanks to the efforts of Van Cleve, Yablo, Chalmers and Hale. Not all explications of the possible senses of conceivability are very satisfying though. This holds in particular for what Chalmers calls ‘positive conceivability’. Positive conceivability seems in some cases to involve forming a mental image of a situation. In these cases it is (more or less) clear what conceiving that p means, but much less clear why our faculty to form such an image should speak for the possibility of the situation so envisaged. It seems to be a question for empirical psychologists to figure out whether or not we are unable to perceptually imagine impossible situations.

Much worse is the fact that in many circumstances no perceptual representation is relevant for the possibility of a situation. Consider Putnam’s Twin Ear or Chalmers’ Zombie World. Perceptually speaking they are both indistinguishable from the actual world (at least that is the idea), but are supposed to be different nevertheless. How can I form a positive representation of such a situation? Here is Chalmers:

In these cases, we do not form a perceptual image that represents S. Nevertheless, we do more than merely suppose that S, or entertain the hypothesis that S. Our relation to S has a mediated, objective character that is analogous to that found in the case of perceptual imaginability. In this case, we have an intuition of (or as of) a world in which S, or at least of (or as of) a situation in which S, where a situation is (roughly) a configuration of objects and properties within a world. We might say that in these cases, one can modal imagine that P. [...] ‘Modal imagination’ is used here as a label for a certain sort of familiar mental act. Like other such categories, it resists straightforward definition. But its phenomenology is familiar: One has a positive intuition of a certain configuration within a world, and takes that configuration to satisfy a certain description. (2003: 151)

If that sounds like a familiar mental act for anybody, then what I am going to say in this paper is not of much concern for such person if she can convince herself that ‘It is modally imaginable that p” implies ‘p is possible’. To me, the notion of positive conceivability, in particular the notion of modal imaginability, is either wholly unfamiliar or identical with perceptual imaginability (and then likely to be uninteresting for matters of modality) or boils down to what Chalmers called – ‘negative conceivability’. To the latter we shall now turn. Again Chalmers:

The central sort of negative conceivability holds that S is negatively conceivable when S is not ruled out a priori, or when there is no (apparent) contradiction in S. [...] And we can say that S is ideally negatively conceivable when it is not a priori that – S. (2003: 149)

‘Ideal’ here means ‘cannot be trumped by better reasoning’ such that a better reasoner with better intellectual capacities could rationally defect the justification for not finding – S a priori. For the purpose of this paper I will – following Peter Menzies – use the notion of an ideal reasoner, presupposing it is coherent. In my terminology, an ideal reasoner is simply a universal Turing machine with infinite storage capacity that is fed with the relevant information and rules of inference. In terms of a Turing machine, to be a priori that – S is that the Turing machine comes up with a proof that – S. This sounds like a straightforward coherent notion. (I will discuss the difference between Chalmers notion of ‘ideal’ and my notion in Appendix I of the paper in Question 1.)

This notion of conceivability has obvious advantages. It seems to be a matter of logical analysis whether or not anything is conceivable, and it seems to suit our discursive practice. If

a philosopher claims to find some S conceivable, his colleagues quickly try to come up with a relevant necessity he might have overlooked or with a hidden inconsistency that makes – S a priori after all. If asked whether zombies are possible I try to find an inconsistency in their supposition, not finding any, I take this to be evidence for their possibility.

Chalmers is rather confident that the following is true:

(NC) Ideal negative conceivability entails possibility.

We will not go much deeper into Chalmers’ discussion at this point. However, he argues for the truth of (NC) by rehearsing a number of purported counterexamples and so-called “inscrutabilities”, i.e. necessary truths undiscoverable a priori that preclude us from recognizing a priori that actually – S and thus lead to modal error. We will comment on his discussion in Question 2 of Appendix I. But in order to clarify (NC), we will explicate it a little more.

In fact, (NC) connects our reasoning abilities (or an idealization thereof) with the domain of possibilities. We can take our reasoning abilities to be modeled by a proof theory – a set of syntactical rules that allow certain inferences, in particular inferring certain possibilities from their negative conceivability. The domain of possibilities can be understood as the semantic counterpart of this proof theory. The claim that ideal negative conceivability entails possibility is the claim that the proof theory is sound, i.e. that if under ideal circumstances

in accordance with the rule of negative conceivability (whatever that is), it is also the case that

Therefore (NC) translates into the following:

(NC') \forall S (\neg \exists p : S is inferred via the rule of negative conceivability under ideal circumstances) \implies \neg \exists S (\neg \exists p : \neg S is inferred via the rule of negative conceivability under ideal circumstances) is valid).

2. DOUBTS ABOUT NC

One of the main problems of negative conceivability is that it seems to be based on an epistemological salto mortale. If I claim that p is possible because it is not a priori that not p, this is clearly an instance of I lack a reason against p, therefore p, which seems spurious:

(DO) I acquire evidence in favor of a proposition’s possibility, by finding myself without evidence against its truth. That would be very strange to say the least. (Yablo 1993: 8)

What I dispute is the contention that if a concept or state of affairs is not logically impossible, then it is ‘logically possible.’ It hardly follows that, because a certain thing cannot be proved to be impossible by a certain method, it is therefore possible at any sense of ‘possible’ whatever. (Van Inwagen 1998: 71)

Bob Hale has recently answered this challenge posed by Yablo and Van Inwagen when defending his necessity based account of modal knowledge.

---

3. WHAT POSSIBILITY IS

Thus NC seems to be a central thesis for Chalmers as well as for Hale. But not only these modal epistemologists are committed to NC, it is also a vital ingredient of the response-dependence-analysis of the modal notions, provided by Peter Menzies.\(^1\)

One way to explain what a notion means is to provide a reductive analysis, reducing the notion to something more familiar.\(^2\) The (infamous) "Blackburn-Dilemma" convinced many that such an analysis is not feasible for modal notions (including natural, moral, logical necessity). If we are to answer the question "Why is A (naturally, morally, logically) necessary?" we could of course answer this question by giving a proof for A from B, which provides an explanation for the necessity of A if we know why B is necessary. But this would anyway only explain the necessity of some particular A, but not necessity as such. If necessity as such is what we are after, a proof of necessity from some other necessity cannot explain the notion as such. But if we explain, on the other hand, necessity in terms of something contingent, necessity will not be explained but rather undermined. How could necessity ever flow from something merely contingent?

Either the explanandum [sic] shares the modal status of the original and leaves us dissatisfied, or it does not, and leaves us equally dissatisfied. (Blackburn 1993: 54)

But assuming that this really is a dilemma, how can we then explain the modal notions? One way to respond to this dilemma is to foreswear reductive analyses and to direct attention to the possession conditions of the notions in question. (In this case, modal concepts and statements are treated as primitive and analyzable, the truthmakers of modal statements are simply modal facts.)

A philosophical explication of modal concepts should then consist in providing the possession conditions of these concepts, which are those "aspects of the practice of subjects possessing the concept that are essential to their competence with the concept" (Menzies 1998, 263). Given the connection philosophers saw between conceivability and possibility, a response-dependence analysis might seem to suggest itself as the proper analysis of these possession conditions. A concept has a response-dependent character if the correct application of the concept implicates a human response in the manner of a secondary quality concept, or, more formally:

\textbf{Definition 3.1}

The concept of a property C is response-dependent if there is some response $R_x$ (sensory, affective, or cognitive) such that the following biconditional holds true a priori: $x \in C \iff x$ is disposed to elicit response $R_x$ in suitable subjects in suitable conditions.

The prime example for a concept that is response dependent in this sense is a colour concept, like RED, if analyzed in the traditional way as a concept of a second order property. RED is then analyzed as

\begin{itemize}
  \item Menzies 1998, see also Fuhrmann 2001 (for another response dependence approach to modality), Wright 1992, Jackson & Pettit 2002 (for response dependence in general).
  \item For an example of a reductive analysis of modality see Lewis 1986.
  \item Blackburn obviously means "explanans".
\end{itemize}
This analysis gives the possession conditions of a concept such as RED insofar as normal subjects take their colour experiences normally as primary criteria for the application of colour concepts, but refrain from doing so, if intertemporal or interpersonal differences occur. In such cases, either some subject or some conditions will be dismissed as "normal" by correcting practices. Thus, instead of reductively analyzing the concept in question, these biconditionals are summaries of specific features of the particular reaction as well as of the correcting practices which must both be implemented in a subject for the subject to possess the concept in question. As an example, subjects which would apply colour terms as a reaction to pain-experiences would not possess colour concepts; neither would subjects which disagree about the actual colour of the wallpaper in a dark room who would rather trace the source of their disagreement to differing preferences than finding it a helpful suggestion to turn the lights on.

Applying a response-dependence analysis to modal concepts presupposes that the possession conditions of modal concepts are relevantly similar to the possession conditions of colour concepts. In Menzies' analysis this similarity should be twofold: there should be a primitive reaction serving as a primary criterion for applying modal concepts (this is supposed to be conceivability) and there should be corrective practices we engage in whereby we refine the responses that count as veridical indicators of modality. However, the primitive reaction turns out to be a complex mental ability (for this point see Question 6 in Appendix I):

What exactly is conceivability? What does the mental ability to conceive something consist in? [...] [The] mental ability to conceive something is really a complex ability, consisting in two abilities to suppose that the state of affairs holds without being able to reduce this supposition to absurdity. Clearly, this complex ability presupposes a number of other, more complex abilities: first, the ability to entertain suppositions; and secondly, the ability to infer other propositions, in particular those propositions from suppositions. (Menzies 1998: 265)

This notion of conceivability is obviously identical to what Chalmers calls 'negative conceivability'. Now, since this "complex ability" is rather demanding of a subject and a response-dependence biconditional has to state the conditions for a veridical application of a concept, Menzies introduces the notion of an ideal concever, a being which does not suffer "any of the limitations discounted by our practice" which are "limitations due to inadequate critical reflection, limitations due to lack of concentration and attention, limitations due to external interference, limitations due to insufficient memory" and more besides. This ideal concever is clearly identical with the ideal reasoner considered above (see also Question 1 of Appendix I). Given this notion, Menzies (1998: 269) suggests the following biconditionals as stating the possession conditions of possibility and necessity:

(3-1) It is possible that $p \rightarrow p$ is conceivable by an ideal reasoner.
(3-2) It is necessary that $p \rightarrow \neg p$ is not conceivable by an ideal reasoner.

We will be concerned with (1) only. Given Menzies analysis of 'conceivability' and our analysis of 'ideal negative conceivability', (1) translates into

(3-1') It is possible that $p \rightarrow \neg p$ is ideally negatively conceivable.

The right to left direction of this biconditional is equivalent with NC and thus implies NC*. If the ideal reasoner is not entitled to infer the possibility of some sentence $p$ because of finding $p$ negatively conceivable, this inference must be found for the biconditional (3-1) to be true. Possibility is being found negatively conceivable by an ideal reasoner (see also Question 7 of Appendix I). We will not go any deeper into Menzies account at this point, for we have already succeeded in what we wanted to show: the modal epistemology supported by David Chalmers, the necessity based account of our modal knowledge endorsed by Bob Hale, and the response-dependence analysis of the modal notions put forward by Peter Menzies all include a commitment to NC*. In the rest of the paper I will show that NC* is false in its intended meaning and that this can be seen on logical considerations alone.

4. MODELLING THE IDEAL REASONER

To show that something is false on logical considerations alone is not easy to achieve if that something does not by itself suggest a particular logic. Any such argument to the effect that some principle $X$ is false in a formal model $M$, invites the reply that $M$ might just not have been the only adequate model. In the case considered here, NC* does not by itself suggest a particular logic. In fact it can be shown that NC* comes out true of some logic. But -- so I will argue -- the logics of which NC* is true are far too weak for the intentions of Chalmers, Hale or Menzies. The logics adequate to the intentions, in terms of their expressive powers, are, however, incompatible with NC*.

We now face the task to translate everything said so far into a formal framework that allows us to prove the falsity of NC* in its intended meaning. We have argued above that NC* is a claim about the soundness of a rule, but which rule? Rules, quite generally, are conceived as a pair consisting of a finite set of preconditions and a conclusion. This, for example, is the rule of disjunction introduction ($V$ Intro) in First Order Logic:

$$\begin{align*}
\therefore & \quad P \lor \ldots \lor P' \lor \ldots \lor P'' \\
\end{align*}$$

The rule states schematic premise formulas which might at some point be derivable in a proof. In this case the schematic premise formula $P_i$ is considered derivable. Now the rule $V$ Intro tells us that if an instance of $P_i$ is derivable in a proof, we can infer an instance of the schematic conclusion formula. In this case we are allowed to infer a disjunction with $P_i$ being one of the disjuncts.

Now what about the rule that NC* is concerned with? We know that it is supposed to infer $p'$'s possibility from its negative conceivability. So we know at least what the conclusion formula must look like:
Definition 4.2 (monotonic axiomatizability)
A logic is called monotonic axiomatizable if it has an axiomatization which consists only of monotonic rules containing only derivability claims.

Since $\Box$ Intro is a nonmonotonic rule, a logic including it will not be monotonic axiomatizable. $\Box$ Intro is the proof-theoretic pendant of negative conceivability; without it we could not infer non-trivial possibilities. A possibility is trivial if the formula in the scope of the possibility operator is valid if necessitated. Note that in normal modallogics non-trivial possibility theorems simply do not occur. If we are interested in a logic of negative conceivability, we need them, however. It is non-trivial possibilities that negative conceivability is all about. To accommodate them in our logic, we will have to depart from normal modal logics in a semantic and a syntactic respect. Semantics first:

In Kripke's A Completeness Theorem in Modal Logic, only those sentences are contained as theorems that are valid with respect to every substructure of the possible worlds. Let us for the moment settle with propositional modal logic. Take a sentence of the form $\Box A$, A being a non-modal formula of $\mathcal{L}_{prop}$ a formal language of propositional modal logic. As we've said, Kripke's semantics will have this sentence as a theorem only if it is valid with respect to every substructure $W \subseteq W$, but then A can be anything but a tautology for it's only those which are guaranteed to be present in every substructure (the proof is given in Appendix II).

Theorem 4.1
For every non-modal formula $A \in \mathcal{L}_{prop}$ and normal propositional modal logic $L$ if $\Box A \in L$, then $\mathcal{C} \models L$, and A is a truth-functional tautology provided it is consistent.

So these semantics are clearly inadequate if we want to model the logic of negative conceivability. To model such logic, or at least a fair portion thereof we need non-trivial possibilities as valid formulas and as theorems, for otherwise what is negatively conceivable if sound tracks what is necessary (and then trivially entails possibility).

So instead of having a variable substructure of $W$ determine validity, we assume a fixed space $W$, containing all possible interpretations of a language and identify interpretations in possible worlds. To show that this works, and to show in what way it departs from standard modal logic, we will give an example in the language of propositional modal logic (following Schurz 2001).

Our logic $\mathcal{C}_{prop}$ consists of a denumerable set $P$ of propositional variables $p, q, ...$, and the standard connectives and operators $\neg, \land, \lor, \rightarrow, \leftrightarrow, \Box, \Diamond$. We assume standard interpretations (truth valuations): $I : P \rightarrow \{0, 1\}$ and a classical semantics for non-modal formulas. As we've said, we interpret propositions in possible worlds, $W$ is the set of all possible worlds (interpretations $I : P \rightarrow \{0, 1\}$). A sentence is called logically true if it is true at all worlds. Now we can add the semantical rules for the truth value of modal sentences: $\mathcal{C} \models L$ (in a given world) iff $A$ is true in all possible worlds iff $A$ is logically true. The truth of necessitated sentences is obviously logically determined, if it is true, it is logically so, if it is false, logically false:

- If $\models \mathcal{C} A$, then $\models L A$
- If $\models \mathcal{C} \neg A$, then $\models L \neg A$ ($\models L \neg \neg A$)

Giving the semantics like this, we immediately have all theorems and rules of SQ, but it also implies that for every propositional variable $p \in P$, $\Box p$ is a theorem of $\mathcal{C}_{prop}$, and the same holds
for every $C_{\text{PROP}}$-consistent sentence $A$. However, this is just as it should, for these are the non-trivial possibilities we wanted from our semantics.

This comes as a surprise though. $C_{\text{PROP}}$ is not closed under homomorphic substitution. This is easy to see. Consider $\neg \varphi \land \varphi$. As we've seen, this follows because $\varphi$ can be made true by at least one of the $\varphi(p^\top \land \neg p^\top)$, which is a substitution instance of $\varphi(p^\top \land \neg p^\top)$. But it is clearly not the case that $\neg \varphi(p^\top \land \neg p^\top)$, since every interpretation makes $p^\top \land \neg p^\top$ false.

So we cannot have an unrestricted substitution rule in our proof theory (if it's sound). However, $C_{\text{PROP}}$ is closed under syntactically isomorphic substitution and semantically isomorphic substitution (a substitution function is semantically isomorphic, i.e., it preserves the semantic freedom of interpretations). It is arguable that this is enough for a proper logic (see Schurz 2001). That $C_{\text{PROP}}$ is not closed under homomorphic substitution is related to the issue of monotonic axiomatizability (not being closed under homomorphic substitution implies not being schematically axiomatizable or not being monotonically axiomatizable; the proofs are given in Schurz 2001).

Theorem 4.2
$C_{\text{PROP}}$ is closed under all syntactically isomorphic and under all semantically isomorphic substitutions.

Theorem 4.3
If a logic is schematically and monotonically axiomatizable, then its theorems are closed under (homomorphic) substitutions.

Since $C_{\text{PROP}}$ will not be closed under homomorphic substitutions, as we have shown, we will not be able to give a monotonic axiomatization or will not be able to keep the axiomatization schematic. Since we want to model a logic that includes the rule $\neg \varphi$, we can keep all rules and axioms schematic. In Appendix II we will give an axiomatization $A_C$ of $C_{\text{PROP}}$ which is schematic, sound, complete and decidable (proof also in the appendix).

Theorem 4.4
The axiomatization $A_C$ of $C_{\text{PROP}}$ is schematic, sound, complete and decidable.

This logic gives us a formalization of negative conceivability. Since $C_{\text{PROP}}$ is sound, $\neg C$ is true in this logic. Every sentence which is negatively conceivable in $C_{\text{PROP}}$ and is therefore inferred to express a possibility does express a possibility, which is good news for Chalmers, Hale, and Menzies. Moreover, $C_{\text{PROP}}$ is complete, thus all possibilities are negatively conceivable, which is good news for Menzies (remember the left to right direction of the biconditional 3-1*). If all works so well for the propositional case, why not consider stronger logics?

We shall call a logic a 'C-modal expansion', $C_L$, of a logical system $L$ with an interpretation semantics if we add to $L$ the two modal operators, $\Box$ and $\neg$, and their respective semantical interpretation clauses. Let's consider a C-modal expansion of First Order Predicate Logic, $C_{\text{FOPL}}$. We would like to know some of the metalogical properties of $C_{\text{FOPL}}$. The first thing we would like to know about our model of negative conceivability is whether it is complete. If it is not complete, that might not be a problem for the epistemologist (like Chalmers or Hale). There might be a computational limit to what we can know a priori or to what we can infer just on the basis of logical considerations alone. But completeness should matter for Menzies. If it is a priori that what is possible is what is negatively conceivable, this logic should be complete or else the left to right direction of the biconditional (3-1*) is false. Now, it is easy to see that the first order variant of our logic of negative conceivability is incomplete, if it is assumed to be the logic by which an ideal reasoner infers possibilities. To make this point as clear as possible, we shall give a very intuitive proof. So let us— for reductio— assume that $C_{\text{FOPL}}$ is algorithmic (its theorems are enumerable by a deterministic Turing machine) and complete.

We will construct a negative converter, HALE 9000, a huge computer that we have programmed with the axioms of First Order Logic (a base class of relevant absolute necessities just in the sense of Hale's necessity based account). We know that this base class is complete, since we know that FOL is complete. Therefore we have ideal conditions in the sense considered above, we have infinite storage capacities, all knowledge, you know, we know all of them and can now start to infer possibilities, check proofs, whatever we like. We like to check for theorems;

HALE 9000 is programmed such that for every formula $A$ that we want to infer, HALE 9000 starts two proofs that it carries out simultaneously. In the first proof it tries to prove $\Box A$, in the second proof it tries $\neg A$. For every formula $A$ of $C_{\text{FOPL}}$, either $\Box A$ or $\neg A$ is valid in $C_{\text{FOPL}}$. Thus, if $C_{\text{FOPL}}$ is complete, HALE 9000 will arrive at a proof for one of them after finite time.

But then we have found a purely mechanical way to decide theoremhood for FOL. First Order Logic is a proper part of $C_{\text{FOPL}}$ and, as we can see in Figure 4-1, checking some formula of FOL for theoremhood should be a matter of finite analysis, just as it is for any other modal formula of $C_{\text{FOPL}}$. But theoremhood for First Order Logic is not decidable, as was proved by Church in 1936. Therefore $C_{\text{FOPL}}$ is not complete, or not a logic that could be used by an ideal reasoner.

**Input-Formula**

$A = \forall x (Fx \supset Gx)$

$\Box \forall x (Fx \supset Gx)$

If $A$ is a theorem, HALE 9000 can prove this formula after finitely many steps, if $C_{\text{FOPL}}$ is algorithmic and complete.

$\neg \forall x (Fx \supset Gx)$

If $A$ is not a theorem, HALE 9000 can prove this formula after finitely many steps, if $C_{\text{FOPL}}$ is algorithmic and complete.

**Figure 4-1**
Theorem 4.5
A C-modal expansion of an undecidable logic \( L \), \( C_k \), is incomplete or not algorithmic.

As we’ve observed already, incompleteness might not be too bad for the modal epistemologist. HALE 9000 might almost know about every logical possibility from just knowing all logical necessities. It might be sufficient if HALE 9000 knows some of them for negative conceivable to be a guide to possibility. NC is anyway not a claim about possibility entailing negative conceivable, but the other way round: it just claims that what we infer to be possible—because we found it negatively conceivable—really is possible. So we can ask the following question: Assumed that \( C_{\text{red}} \) is incomplete, is it sound to infer possibilities with the rule \( \Diamond \text{Intro} \) or necessities by way of \( \Box \text{Intro?} \) Theorem 4.5 left us the result that \( C_{\text{red}} \) is neither the logic of an ideal reasoner or incomplete. Therefore it might still be sound, even if we grant that it is incomplete. But again we can prove that soundness cannot hold if we assume the rules for the modal operators to be admissible in \( C_{\text{red}} \). Intuitively, incompleteness means that there is some \( A \) such that \( \neg \Box A \) but \( \Box \neg A \). By \( \Box \text{Intro} \), \( \neg \Box A \) allows us to infer \( \neg \Diamond \neg A \), which is equivalent to \( \neg \Box A \), hence \( \neg \Box A \). Since \( \neg \Box A \), the semantics of the modal operators tell us that \( \neg \Box A \). Therefore incompleteness would lead to an unsound logic if we keep the rules. Q.E.D.

At the same time, unsoundness implies incompleteness (if \( C_{\text{red}} \) is consistent). Assume unsoundness but completeness, i.e. there is a formula \( A \) such that \( 
\neg \Box A \), but \( \Box \neg A \). By \( \Box \text{Intro} \), \( \neg \Box A \), but by the semantics of the modal operators, \( \neg \Box A \), which is equivalent to \( \neg \Box A \). By completeness, \( \neg \Box A \). Q.E.D.

These results can be generalized (proof in Schurz 2001):

Theorem 4.6
For every consistent axiomatization of a logic of kind \( C_k \) in which the rules \( \Box \text{Intro} \) and \( \Diamond \text{Intro} \) are admissible, the following holds: \( C_k \) is semantically sound if it is semantically complete.

Therefore, as soon as we are dealing with the modal status of sentences that stand in inferential relations, such that a logic with the expressive powers of First Order Logic is needed to adequately represent the intuitively obtaining logical relations, negative conceivable will not entail possibility.

If First Order Logic (or some other undecidable system) properly captures the notion of logical truth (and thus of logical necessity), then knowledge of logical necessity does not lead to knowledge of logical possibility.

If First Order Logic plus meaning postulates properly capture the notion of analytic truth (and thus of conceptual necessity), then knowledge of conceptual necessities does not lead to knowledge of conceptual possibilities, etc.

In all these cases, negative conceivable does not entail possibility.

5. CONCLUSIONS

We have started with the observation that some modal epistemologists seem to argue in favour of the thesis that knowing what is (logically) possible is achieved by an inference from the fact that something cannot be shown to be (logically) impossible. Since it is knowledge which is at issue, the question is whether such inferences are justified. The latter involves that these inferences track truth under ideal conditions, or in other words, that ideal negative conceivable entails possibility.

We have noted that this entailment is really a claim about the soundness of an inference rule. In our modal reasoning, we draw inferences from beliefs we have. Sometimes we come to believe that a certain state of affairs is possible because we did not find anything among our beliefs that could rule this state of affairs out, we could not show that the state of affairs in question is impossible. True, a lot of things can and do go wrong. We might not know of a relevant necessity that rules this state of affairs out, we might know of it but might not have paid enough attention, we might have confused something, etc. However, the fact that our notion of possibility is such that what is possible is what is not excluded by the laws of logic, language, metaphysics, or physics, it is highly plausible also to assume that this must be the way we get to know of these possibilities. At least under ideal circumstances, when we know all necessities relevant and really concentrate on the issue, negative conceivable should track truth. In this case the inferential going-ons in our minds should coincide with the ontological going-ons. If our mind is in this respect like a proof-theory, drawing inferences by certain rules, then an idealized version of it should be in accordance with what really is possible. If we reason by negative conceivable, this reasoning should be sound, negative conceivable should entail possibility. This seems to be the message of Hale, seems to be defended by Chalmers and is clearly stated by Menzies. As we have seen, this does not hold.

It only holds if we restrict our reconstruction of modal reasoning to a fragment of it, a fragment that has a decidable underlying logic. For Menzies’ analysis this is far from being an option. Of course, there are logical impossibilities not excluded by the laws of a decidable fragment of First Order Logic. If this is a priori, then Menzies analysis will fail. Moreover, such a move seems absolutely independent of motivation. Why should our modal reasoning be so limited?

The results obtained are general, we have given a model of modal reasoning which follows the most natural assumptions about logical necessity, possibility and negative conceivable.

Negative conceivable does not fail to entail possibility because of the existence of a posteriori necessary truths or the existence of weird mathematical truths which are strongly incomputable or anything like that, but fails because it is based on a misconception of what is involved in inferring p’s possibility from not being able to rule it out. What was not noticed is that ‘not being able to rule it out’ is really a claim of non-derivability and that such claims are not reliable as soon as we are dealing with something slightly complex.

REFERENCES


APPENDIX 1: QUESTIONS AND ANSWERS

In this appendix I try to answer to some of the questions that were either raised during discussions of my argument or questions that have troubled me (some, if not most, still do). Suggestions for better replies are highly appreciated.

**Question 1**

Chalmers’ notion of ‘ideal reasoning’ differs somewhat from yours. Shouldn’t that make a difference?

**Answer**

No, I think it should not. Chalmers’ notion is this:

[...]

[...] We can say that S is ideally conceivable when there is a possible subject for whom S is prima facie conceivable, with justification that is undefeated by better reasoning. (Chalmers 2003: 148)

Now it might seem as if such an ideal reasoner were never in a position to detach mistakenly by ◊ Intro, say ◊ ¬A, without ceasing to be ideal at the same time, for — by completeness of the underlying logic — longer reasoning would have revealed the impossibility in question and thus undermined the justification for ◊ A.

Note that we are faced with an infinite hierarchy of ideal reasoners such that at any point, given the fact that our inference rule is not sound, we are unjustified to infer the possibility of S from the non-apriority that ◊ S. But then the entailment thesis is empty: if no detachment is ever justified, possibilities will not be accessible by the method of negative conceivability. The thesis that ideal negative conceivability entails possibility will be vacuously true, because the antecedent of: If it is ideally negatively conceivable that S, then S is possible, will not be true for any S. We know a priori that ◊ Intro is not a sound inference rule. This is different from believing that we have not overlooked a relevant necessity or made a mistake in our proof — we could subjectively be justified in believing this, and this justification could be undermined by better reasoning. But there is no justification for using an a priori unreliable rule of inference and thus never a justification for detachment that could not be undermined by better reasoning. But see also Question 3.

**Question 2**

It seems that in his paper (Chalmers 2002) Chalmers already went through all possible counterexamples and dismissed them. How come that you find a flaw? Shouldn’t this already been among his “inscrutabilities”?

**Answer**

Chalmers seems only to be concerned with the question whether we can know all truths a priori if we are given a full qualitative description of the world. He thinks we can and I did not intend to challenge this (see also Question 3). What I think he overlooks is that even if there are no inscrutabilities in the sense that there are truths that we cannot know a priori on the basis of a complete qualitative description of the world (plus some indexical knowledge), negative conceivability is still no guide to possibility. To rule out inscrutabilities is not enough.

**Question 3**

(Chalmers) It seems that your argument turns on replacing ‘negative conceivability’ with a notion defined in terms of formal provability. Certainly the thesis that such notions are equivalent is to be rejected. There is a very good reason to think that ideal reasoning can’t be fully formalized by any logical system — e.g., the reasons tied to Gödel’s theorem. (Every statement of arithmetic is a priori in the ideal sense, but there’s no formal system that proves all of them.) We already know that if one equates ‘S is conceivable’ with ‘S is provable in formal system F’ then the conceivable-possibility link is false — a counterexample will be provided by the negation of a Gödel-sentence for F.

**Answer**

I think this reply rests on a completely useless notion of ‘a priori’. One problem seems to be that we couldn’t know a priori that there are truths we cannot know a priori, if this notion were correct. But to me the lesson of Gödel’s theorem is exactly that we can know a priori that there are truths we cannot know a priori. Anyway, if “there is very good reason to think that ideal reasoning can’t be fully formalized by any logical system” then good bye to the Church-Turing thesis, good bye to logic as the enterprise of explicating the rules of valid reasoning, good bye to analytic epistemology.

However, my intention was to steer clear of second order problems (and inscrutable mathematical truths) by trying to show that even if we do not know a priori that our knowledge of necessities is relevantly incomplete (which should keep us from taking negative conceivability
as an indicator of possibility in any case), but know that we can prove every necessity in finite time, negative conceivability would still not entail possibility.

**Question 4**

Why should Hale be committed to NC? Didn’t you merely show that the inference rule is fallible, something which Hale admitted anyway?

**Answer**

Although Hale is very liberal when characterizing the formal properties our method has to learn what is possible must have, it seems that even if we grant that it might be fallible in all cases in which we didn’t have enough information or made a mistake, it should nevertheless track truth under ideal circumstances. I think this is partly due to the fact that he has the burden of the argument to show that necessity-based methods are sufficient to account for our modal knowledge.

Hale might argue that the errors we make by applying \( \Box \text{Intro} \) are, by the completeness of the underlying logic, corrigible after finite time, and that therefore the situation is not unlike the situation envisaged by falsificationists when they explain our knowledge of *actualia*. (In this case Hale, too, could try to find support in Chalmers’s notion of ideal reasoning.) But the situation is relevantly unlike falsificationism. Falsificationism allows for justified beliefs because (i) we can explain all mistaken beliefs by reference to cognitive limitations (we made a mistake when deducing empirical consequences, we made a mistake when observing the experiment) or insufficient information (this consequence was not yet tested and turns out to be empirically false), and (ii) we can systematically reduce these mistakes (we use microscopes and computers, and test other empirical consequences). Both do not apply here. In our model no information is hidden from us, and we assumed a Universal Turing machine to carry out the proofs. Reasoning by \( \Box \text{Intro} \) does not lead to justified beliefs, because our epistemic situation cannot be systematically improved any more.

**Question 5**

But why should Hale be committed to \( \Box \text{Intro} \)? Isn’t it possible to find an inference rule that would not lead into problems? What about adding a provability operator \( \Box \) to the logic and a modified rule like this:

\[
\begin{align*}
\vdash & \Box P_i \\
\vdots & \\
\vdots & \\
\vdots & \\
\vdots & \Box \neg P_i & \text{conclusion formula}
\end{align*}
\]

Why should that lead to problems?

**Answer**

Even if we ignore the fact that this “new” rule is somewhat artificial in this context and its details are not as easily spelled out as it might seem (what are the introduction and elimination rules for \( \Box \)?) the result would not be different in any way. The accessible possibilities were still restricted to the largest decidable fragment of FOL, and my question was “Why think that modal knowledge is so restricted?”

---

**Question 6**

Couldn’t Menzies just give up the explication of conceivability in terms of provability and treat it as a primitive reaction after all?

**Answer**

This seems to me to be a possible strategy only in cases in which we are sufficiently familiar with the capacity in question. ‘Conceivability’ is notoriously ambiguous so Menzies has to give us some explication otherwise he would characterize the reaction in a “whatever it takes way” (‘ideal conceivability’ is conceivability that entails possibility) which renders the biconditional uninformative. There might be ways to explicate conceivability suitably, but negative conceivability is not among them.

**Question 7**

What about biting the bullet and admitting frankly that there are only those possibilities that are accessible given a decidable fragment of our language?

**Answer**

Given that the modal notions are interdefinable and given that we still want to say that FOL plus meaning postulates is complete, this doesn’t seem to be a way out of trouble. Part of the language would become logically indeterminate with respect to its modal status (it seems to me). In any case this move is so implausible that it would rather undermine the whole project of response-dependence analyses.

---

**APPENDIX II: TECHNICAL DETAILS**

As some might have noticed already, the logic of negative conceivability is based on Rudolf Carnap’s modal logic.¹ The metalogical properties of this logic have been studied by a number of logicians, like Gottlob, Hendry/Prokierka, Makinson and Schurz.² Although these authors have largely ignored the case of a C-modal expansion of an undecidable logic, not much of what follows will be really new (you can find all the proofs for example in Schurz 2001). However, this appendix might be helpful to readers who want to understand the theorems given in the main text.

Before we turn to Carnap’s modal logic and its properties we will briefly prove a result concerning normal modal logics that we have mentioned in the main text in theorem 4.1 (note that in this proof we distinguish between worlds and interpretations in the well-known way, which we don’t have to in the rest of the paper).

**Theorem 4.1**

For every nonmodal formula \( A \in \mathcal{L}_{\text{prop}} \) and normal propositional modal logic \( L \): if \( \Box A \models L \), then:

\( \Box A \models L \) and \( A \) is a truthfunctional tautology provided \( L \) is consistent.

¹ Carnap 1942, 1947.
Proof

If L is the inconsistent logic, the claim is obvious. Assume L is consistent. It is a well-known fact of frame semantics (i.e., semantics with possible world sets plus accessibility relation) that every consistent normal modal logic is either valid on the reflexive singleton frame \([\{w\}, \{w, w\}\})

or on the irreflexive singleton frame \([\{w\}, \varnothing\})

So \(\Delta \models w\) implies that \(\Delta \models w\) on one of these singleton frames (hence true for every interpretation associated with the world \(w\)). Since \(L\) is non-modal, this implies that \(L\) is verified by every interpretation. Hence, \(L\) must be a truth-functional tautology, whence \(\Delta \models L\). Q.E.D.

Now we turn to the specific properties of Carnap's modal logic, the promised axiomatization of a Carnapian propositional modal logic and finally to its metalogical properties, 4-4.

Axiomatization \(A_c\) of \(C_{prop}\)

A simultaneous axiomatization of \(C_{prop}\) can be given by using rules of natural deduction which hold in both directions and reduce complexity in one. For simplicity we assume only two logical connectives \(\land\) and \(\lor\). \(\land\) stands for a literal, a negated or unnegated propositional variable.

We will first state the derivability-rules (\(\vdash\)), then the rules of non-derivability (\(\dashv\)):

\[
\begin{align*}
\text{Lj}, \ldots, L_n \vdash_{C_{prop}} L &\text{ if } L \in \{L_1, \ldots, L_n\} \text{ or } L_1, \ldots, L_n \vdash_{C_{prop}} L \quad \text{(Literal-Conc)} \\
\text{Lj}, \ldots, L_n \vdash_{C_{prop}} L &\text{ if } \exists L' \in \{L_1, \ldots, L_n\}; L' = \neg L' \quad \text{(Literal-Prem)} \\
\Gamma \vdash_{C_{prop}} \neg A &\text{ if } \Gamma, A \vdash_{C_{prop}} \bot \quad \text{(Negation-Conc)} \\
\Gamma \vdash_{C_{prop}} \neg A &\text{ if } \Gamma, A \vdash_{C_{prop}} \bot \quad \text{(Negation-Prem)} \\
\Gamma \vdash_{C_{prop}} A \land B &\text{ if } \Gamma \vdash_{C_{prop}} A \text{ and } \Gamma \vdash_{C_{prop}} B \quad \text{(Conjunction-Conc)} \\
\Gamma \vdash_{C_{prop}} A \land B &\text{ if } \Gamma \vdash_{C_{prop}} A \text{ and } \Gamma \vdash_{C_{prop}} B \quad \text{(Conjunction-Prem)} \\
\Gamma \vdash_{C_{prop}} \neg (A \land B) &\text{ if } \Gamma \vdash_{C_{prop}} \neg A \text{ and } \Gamma \vdash_{C_{prop}} \neg B \quad \text{(NegConjunction-Conc)} \\
\Gamma \vdash_{C_{prop}} \neg (A \land B) &\text{ if } \Gamma \vdash_{C_{prop}} \neg A \text{ and } \Gamma \vdash_{C_{prop}} \neg B \quad \text{(NegConjunction-Prem)} \\
\Gamma \vdash_{C_{prop}} \neg A \text{ if } \Gamma \vdash_{C_{prop}} A \text{ or } \Gamma \vdash_{C_{prop}} \bot \quad \text{(Necissity-Conc)} \\
\Gamma \vdash_{C_{prop}} \neg A \text{ if } \Gamma \vdash_{C_{prop}} A \text{ or } \Gamma \vdash_{C_{prop}} \bot \quad \text{(Necissity-Prem)} \\
\Gamma \vdash_{C_{prop}} \neg A \text{ if } \Gamma \vdash_{C_{prop}} A \text{ or } \Gamma \vdash_{C_{prop}} \bot \quad \text{(NegNecessity-Conc)} \\
\Gamma \vdash_{C_{prop}} \neg A \text{ if } \Gamma \vdash_{C_{prop}} A \text{ or } \Gamma \vdash_{C_{prop}} \bot \quad \text{(NegNecessity-Prem)} \\
\end{align*}
\]

\[
\begin{align*}
\text{Lj}, \ldots, L_n \not\vdash_{C_{prop}} L &\text{ if } L \not\in \{L_1, \ldots, L_n\} \text{ and } L_1, \ldots, L_n \not\vdash_{C_{prop}} L \quad \text{(Literal-Disc)} \\
\text{Lj}, \ldots, L_n \not\vdash_{C_{prop}} L &\text{ if } \neg \exists L' \in \{L_1, \ldots, L_n\}; L' = \neg L' \quad \text{(Literal-Prem)} \\
\Gamma \not\vdash_{C_{prop}} \neg A &\text{ if } \Gamma, A \not\vdash_{C_{prop}} \bot \quad \text{(Negation-Conc)} \\
\Gamma \not\vdash_{C_{prop}} \neg A &\text{ if } \Gamma, A \not\vdash_{C_{prop}} \bot \quad \text{(Negation-Prem)} \\
\Gamma \not\vdash_{C_{prop}} A \lor B &\text{ if } \Gamma \not\vdash_{C_{prop}} A \lor B \quad \text{(Disjunction-Conc)} \\
\Gamma \not\vdash_{C_{prop}} A \lor B &\text{ if } \Gamma \not\vdash_{C_{prop}} A \lor B \quad \text{(Disjunction-Prem)} \\
\Gamma \not\vdash_{C_{prop}} \neg (A \lor B) &\text{ if } \Gamma \not\vdash_{C_{prop}} \neg A \lor B \quad \text{(NegDisjunction-Conc)} \\
\Gamma \not\vdash_{C_{prop}} \neg (A \lor B) &\text{ if } \Gamma \not\vdash_{C_{prop}} \neg A \lor B \quad \text{(NegDisjunction-Prem)} \\
\Gamma \not\vdash_{C_{prop}} \neg A \text{ if } \Gamma \not\vdash_{C_{prop}} A \text{ or } \Gamma \not\vdash_{C_{prop}} \bot \quad \text{(Necissity-Conc)} \\
\Gamma \not\vdash_{C_{prop}} \neg A \text{ if } \Gamma \not\vdash_{C_{prop}} A \text{ or } \Gamma \not\vdash_{C_{prop}} \bot \quad \text{(Necissity-Prem)} \\
\Gamma \not\vdash_{C_{prop}} \neg A \text{ if } \Gamma \not\vdash_{C_{prop}} A \text{ or } \Gamma \not\vdash_{C_{prop}} \bot \quad \text{(NegNecessity-Conc)} \\
\Gamma \not\vdash_{C_{prop}} \neg A \text{ if } \Gamma \not\vdash_{C_{prop}} A \text{ or } \Gamma \not\vdash_{C_{prop}} \bot \quad \text{(NegNecessity-Prem)} \\
\end{align*}
\]

Theorem 4-4

The axiomatization \(A_c\) of \(C_{prop}\) is schematic, sound, complete and decidable.

Proof

Schemacity follows from the very formulation of the rules. Decidability holds, because the rules are recursive (reduce complexity) in the left-to-right direction. So, for each sequent which figures as proof-goal, the search for a proof will halt after finitely many steps at a sequent containing only literals, which can be decided by the start rules. Soundness is shown by simple semantical means. Completeness follows from the fact that the rules are semantically valid in both directions. So, if a sequent is not derivable, its terminating sequent, which contains only literals, is not an instance of a start rule and hence not valid; whence (by induction on the length of the failed proof-attack) also the original sequent cannot be valid. Q.E.D.