DANIEL COHNITZ [*]

MODAL SKEPTICISM.
PHILOSOPHICAL THOUGHT EXPERIMENTS AND MODAL EPISTEMOLOGY

INTRODUCTION

In this paper I examine some of the criticism the method of thought experimentation was confronted with during the last decades, and shall try to defend it – at least for certain problems, restricted by certain constraints. First of all, in section (1) I give a *motivation* to defend thought experimentation against the "skeptic." [1] Confronted with serious criticism, it might be easier to give up that method and to do philosophy without those scenarios. I shall suggest (although not argue for it here) that it would not be so easy to give up this method, and that without it we could no longer answer certain interesting questions. To clarify what I mean by "thought experiment", I shall give a brief *regimentation* of the subclass of philosophical arguments on which I want to concentrate. In (2) I shall discuss one line of attack against thought experiments: *modal skepticism*. Against this attack I shall present in (3) an informational account of possibility that shows how increase in knowledge of *actualia* will result in increase in knowledge of *possibilia*. In (4) I shall briefly address the sources of modal errors. In the concluding part (5) of this paper, I shall give an assessment of what else would have to be done to provide a satisfying methodological account of thought experimentation.

1. THIS WORLD IS NOT ENOUGH

One of the most basic methods of philosophy is and has always been the consideration of counterfactual cases and imaginary scenarios. One purpose of doing so, obviously, is to test our theories against such counterfactual cases. [2]

Although this method is widespread, it is far from being commonly accepted. Especially during the last two decades, it has been confronted with criticism ranging from complete dismissal [3] to denying only its critical powers [4] to a cautious defense [5] of the use of thought experiments as counter-examples. [6]

Of course, one can find innumerable thought experiments which do not constitute clear-cut counter-examples against the theory they are meant to criticize.

281

They quite often appear to be question-begging or indecisive, or it turns out that the intuitions the thought experimenter intends to invoke are not universally shared by his audience. Some might find a certain scenario perfectly conceivable, whereas others are convinced that the case described is definitely impossible, and still others might claim that they do not know what to say at all about such a counterfactual situation.

Diagnosing such disagreement leaves us with two possible reactions: we could simply abandon the method and do philosophy without it, or we could try to find out whether the method can be trusted (at least) under certain circumstances, meeting certain constraints. The latter reaction would involve an epistemological inquiry. What I suggest (although I cannot argue for it here) is that, in the case of thought experimentation we lack an alternative method that serves the same purpose. Roy Sorensen, Sören Häggqvist, Frank Jackson[7], and others have argued that we could no longer answer certain interesting questions without it.[8] So what we are left with (if Sorensen, Häggqvist and Jackson are right) is the task of finding the sources of disagreement about certain thought experiments, and proving to the *skeptic* that an inquiry concerning modal epistemology is a promising enterprise.

Necessity Refuters – a Description

But let me first characterize in a semi-formal way what I have in mind when I speak of 'philosophical thought experiments' or 'modal arguments'. As Sorensen and Häggqvist have argued, there is an important subclass of arguments usually labeled 'thought experiments'[9] (Sorensen calls them 'necessity refuters') which consist of the following ingredients[10]:

(I)   The target claim T: This can be any statement which implies a necessity claim. It might be e.g. some reductionist thesis (in the sense that to reduce either means to claim identity between two rigid designators or involves some strong supervenience), a conceptual analysis (stating necessary and/or sufficient conditions for the application of a certain concept, claimed to be conceptually necessary), a law statement and so forth. In fact, every deductive argument as a whole could do, since it implies an entailment claim (that the premises logically entail the consequent). T itself or its implied consequence is of the form $\Box(A \rightarrow B)$.

(II)   The accommodation of refuter R: The partial description of a possible world that (i) belongs to the set of possible worlds (I) is talking about, and (ii) in which the refuter R, $(A \& \neg B)$, is true.

(III)   The possibility claim $\Diamond(A \& \neg B)$, following from (II).

(IV)   The conclusion of the argument of the form $\neg\Box(A \rightarrow B)$, or some inference leading to the negation of T, the refutation of the target claim.

If one wants to find examples of such a necessity refuter, one could turn to the philosophy of mind: Let a version of physicalism be the target claim *T*. Physicalism is usually taken to hold that mental truths are reducible to physical truths and, hence, that the so described mental and physical events are identical. This reduction thesis implies that identity is a metaphysical necessity in the Kripkean sense. It follows that if physicalism is true, it is metaphysically necessary that if all physical facts are given, all mental truths are fixed as well. This establishes step (I).

Now consider the metaphysical possibility of a zombie, who is physically indiscernible from us, but lacks all *qualia*, i.e., phenomenal states. And let a "zombie world" be a world physically indiscernible from ours but inhabited by zombies instead of human beings. Let's assume that one could accommodate such a situation as a possible world in the relevant metaphysical sense[11], and that the phenomenal states the zombie world is lacking are mental facts in our world. This would establish (II) and hence (III), namely that there is a metaphysically possible world in which all statements about physical events are true as well, but the relevant mental statements are not. From here we could go on to the refutation of the target claim and conclude that physicalism is false.

2. MODAL SKEPTICISM

Arguments of this sort are certainly formally valid. What is doubtful about such arguments is the justification of the modal premise (II): can we establish that a certain state of affairs is possible in the relevant sense? Skepticism with respect to thought experiments often amounts to giving a general negative answer to this question. One of the strongest criticisms of the method of thought experimentation, especially of arguments involving modal premises (like instantiations of the schema given above), is "modal skepticism" as explicated and defended by Peter van Inwagen.[12] There are, of course, other valid modal arguments which similarly involve an assertion of possibility as their crucial premise. In fact, Peter van Inwagen characterizes a subclass of such thought experiments which prima facie differs in structure from our regimentation of a necessity refuter.[13] He is concerned with arguments of the following form:

It is possible that I exist and nothing material exists

Whatever is material is essentially material

hence,

I am not an immaterial thing.

Given the example of an argument against physicalism above, it is easy to see that Inwagen's "modal arguments" could be structured in the form of necessity refuters – if that would matter. However, the epistemological problem that

concerns the justification of premise (II) is exactly the same problem in all arguments of concern to Inwagen.

Inwagen starts with the observation that one can always turn a modal argument into a "contrapositive" argument – which is to say that whenever somebody argues from a refuter R to the negation of T, the strategy to argue from T to the negation of R is always an open choice. It is not entirely clear what the argumentative status of Inwagen's observation is. If he intended it to reveal a problem specific to thought experiments, it must be noted that *this* is not a specialty of modal arguments: in fact every deductive argument can be countered by a contrapositive argument. This triviality instantiates the logical structure of the so-called "Duhem-Quine Thesis": if an empirical theory and a set of background assumptions are inconsistent with certain empirical data, logic alone can't decide between a refutation of the theory, the background assumptions or the data. This correct logical observation applies to non-modal arguments as well.

The only difference with non-modal arguments is that we possess a philosophical theory about ways of judging the reliability of our premises relative to the negation of our conclusion in their case. Hence we can explain why not every contrapositive argument appears to be as convincing as the "original." This philosophical theory consists in an elaborate epistemology which tells us under what conditions we can trust the empirical data more than our empirical theory.[14] In the modal case, we are lacking such a straightforward epistemology. Peter van Inwagen gives an outline of a rather weak modal epistemology (based on an idea of Stephen Yablo) and concludes that this epistemology should lead us to a general skepticism concerning all modal claims which are not bound to our everyday life. Whoever claims to have a special insight into possibilities beyond that is, according to Inwagen, simply fooling himself:

The illusory character of their conviction [to have a capacity to determine the truth value of such possibilities] is sometimes disguised by talk of "logical possibility," for it is often supposed that there is a species of possibility that goes by this name and that one can determine a priori whether a concept or state of affairs is logically possible. But there is no such thing as logical possibility – not, at least, if it is really supposed to be a species of possibility.[15]

Nevertheless, and this is even admitted by Inwagen, we *know* of certain modal truths. We know that certain propositions are possible, although we do not know that they are true (maybe because they are false, or because we don't know whether they are true or false). This knowledge is of possibilities which are "bound to our everyday life." Unfortunately, Inwagen doesn't provide any better definition. Maybe we can circumscribe *basic modal knowledge* as knowledge of possibilities we are most likely to consider to rank our everyday preferences. Inwagen claims that these modal claims of ours are not as doubtful and indeed may represent modal knowledge. But if $\Diamond p$ can not be inferred from the fact that we could not prove p to be impossible, and is not inferred from p's being true,

we run into an epistemological problem concerning such basic modal knowledge – a problem which, according to Inwagen, we obviously have: how do we acquire basic modal knowledge?

Although Inwagen claims not to know how to answer questions like that[16], he suggests that Stephen Yablo's modal epistemology[17] is a step towards a solution. Yablo defends a weaker version of *Hume's maxim*,

$\Diamond p$ iff p is conceivable.

Yablo's weaker version restricts 'p is conceivable' to 'I can imagine a possible world I take to verify p', and defends the thesis that

If I am able to imagine a world I take to verify p, I am *prima facie* justified in believing that $\Diamond p$.

Yablo's thesis is weaker than Hume's, since not all possible ways of conceiving something are said to be reliable guides to possibilities. Only this certain sort, which involves the objectual (instead of a mere "propositional") imagination of a world I take to verify the proposition in question, can be defended as the methodologically useful and epistemically successful way of conceiving something. Inwagen accepts Yablo's thesis, since it seems that this constraint is strong enough to rule out all apparent possibilities which go far beyond our everyday life. Any possible world to verify such a p would be either too difficult to specify in sufficient detail, or simply no world that could be taken to verify p (for apparently all cases which are usually considered in necessity refuters).[18] For example, to specify a possible world which could be taken to verify

(P) There is a naturally purple cow.

would be to spell out in detail "that there is a possible purple pigment such that the coding for the structures that would be responsible for its production and its proper placement in a cow's coat could be coherently inserted into any DNA that was really cow DNA – or even "cow-like-thing-but-for-color" DNA."[19] Whereas imagining the headline 'Mutant purple cows cloned in New Jersey lab' on the cover of a conceivable issue of *Scientific American*, would surely not suffice to verify (P), since this world could only be taken to verify a disjunctive proposition with 'It's the April 1 issue of *Scientific American* and it's a hoax,' as one of its disjuncts besides (P).

On the other hand, Yablo's thesis seems to justify our unquestioned basic modal knowledge, since for these cases it seems relatively easy to imagine in considerable detail a world verifying p. For example, a proposition like

(Q) The desk could have been two feet to the left.

is surely basic modal knowledge in Inwagen's sense. For (Q), it wouldn't be too hard to commit oneself to the possibility of a whole, coherent reality of which the truth of

(Q*) The desk is two feet to the left.

is an integral part. We simply would alter the history of the actual world insofar that, e.g., I decided differently when furnishing my office, and placed the desk two feet to the left. This account would rule out weird possibilities, whereas it would explain our basic modal knowledge. Hence, Yablo's modal epistemology supports modal skepticism.

[N]othing we imagine is absolutely impossible.
David Hume

3. Basic vs. Remote Modal Knowledge

Well, actually it does not. (P) and (Q*) would both be ruled out. Why should we think that a world with a different decision made by me can be specified coherently? Because we did that? Not quite! One of the differences between (P) and (Q*) is not that (P) is so far away from the actual world that we would have to fill in too many unknown details in the design of a possible world. The truth is that in (P)'s case, we know too well what details there are to be filled in (about pigments and cow DNA). It is more likely that the amount of biochemical knowledge we already have simply rules out (P).[20] Whereas in (Q*)'s case it is not as likely that our background knowledge might rule (Q*) out. So again, why should we think that a world with a different decision made by me can be particularized coherently?

We can think so since there is no *available information* ruling it out. For all we know (better: for all we justifiably hold to be true), there is no inconsistency in altering the history of the world in that way. Of course, we don't know too many of the natural laws which govern my mind, and which determined my "choice" when I furnished my office. Maybe there are no determining laws. And if there are such laws, maybe they are compatible with a different choice by me. We simply don't know. So even worse for (Q*)? It's not only that details are missing (as in (P)'s case) – we even don't know what kind of details are to be filled in. But as long as we don't know that, why should there be any reason not to be justified in believing that it is possible that (Q*) or that (P)? What is established so far is that we do not know these statements with *certainty*. We might have made an error, we might have overlooked an impossibility. Is this a reason not to be justified in believing p, if you are not certain that p?

So far, we have established that the *distinction* modal skepticism wants to draw between basic and remote modal knowledge is absolutely unmotivated by Yablo's epistemology. But Inwagen's claim that "anybody who accepts [Yablo's] account should be a modal skeptic" is not fully refuted yet. What we now have to show is that modal skepticism does not follow from Yablo's thesis. This modal skepticism would now be general in kind, since we have just seen that a basic/remote distinction is definitely not supported.

To argue that we are sometimes justified in believing something to be possible just *because* we don't know it to be impossible, is to reject Inwagen's main claim that what is not known to be impossible isn't therefore possible. So the task now is to show that such inferences can indeed yield justified beliefs. One way to do that is with an *informational account of possibility*. In what follows, I shall sketch such an account, which is influenced by Jon Barwise's paper "Information and Impossibilities." [21] The purpose of this excursion is to give a partial explanation of some of the notions just used. Most other accounts of thought experiments fall short of providing such a further explanation of what they actually mean when they demand that the setting of a thought experiment should be "relevant", or how modal knowledge can increase. I hope that the following will be somewhat clarifying here.

When one tries to use the notion of possibility to explicate *information content* (like, e.g., Dretske's information-theoretical account of knowledge[22]), one easily arrives at the inverse relationship principle which is the center of *informationalism*:

> *The Inverse Relationship Principle* (IRP): Whenever there is an increase in available information there is a corresponding decrease in possibilities, and vice versa.[23]

A problem of such accounts is that they usually are not able to provide a sufficiently fine-grained analysis. Consider some exercise in a math textbook: you are asked to compute a number, say the product of two large numbers. After reading the instruction, you simply do not *know* what the solution (the product) is. You start computing and come to the solution. What is this new knowledge you have when you now know the value of the product? What is the content of this new information?

According to the IRP and a standard interpretation of impossibilities, the information content would be analyzed in terms of possibilities which are now ruled out by the computation. But since the solution of this mathematical equation is mathematically necessary, there simply are no such possibilities. The set of possibilities that are ruled out by the computation is empty. Hence, the new information can't be modeled within such a framework.

One way to overcome this problem is to use impossible worlds as the members of the set of worlds now ruled out by the computation. The set would not be empty anymore and would allow for differentiation between knowledge of

different necessary truths. The idea behind this version of informationalism is "that impossibilities are those states of the system under investigation that are ruled out by (i.e., compatible with) the currently available information about the system. States not so ruled out are possibilities."[24] This obviously supports a relativistic and pragmatic notion of possibility. The set of possible worlds (or "possible states", if one wants to emphasize the link between informationalism and moderate modal realism) is (i) relative to a knowledge situation ("currently") and (ii) relative to the availability of information about the relevant issues. Condition (i) makes perfect sense when one wants to talk about possibilities and their inverse relationship to information. Relative to newly available information, certain states of the system under investigation can turn out to be impossible states during the inquiry. But before this information was available, they were possibilities – that is why we now can describe the different knowledge situations. In fact, our modal knowledge did increase.

This positive account of how we can increase our knowledge of possibilities can be accompanied by another argument against modal skepticism. In this paragraph I shall show that modal skepticism is ill-motivated if it is supposed to be a *local* skepticism, a skepticism that is not supposed to shed equal doubt on all of our beliefs. Consider Peter van Inwagen's example of the infallible Standard Atlas:

Suppose that the infallible Standard Atlas marks many islands as uninhabitable, none as inhabitable, and makes no claim to completeness in that matter. We could, if we liked, say that the islands marked 'uninhabitable' in the Standard Atlas were "cartographically uninhabitable." In doing this, we should be calling attention to the fact that our knowledge that these islands were uninhabitable had a certain source. But would there then be any sense in saying that an island was "cartographically inhabitable" just in case it was not cartographically uninhabitable? Very little; I should think.[25]

Inwagen suggests that if we only know that some islands are uninhabitable, we aren't *therefore* justified in believing that it is possible to live on the others. The argument runs like this:

(P1) We have some negative modal knowledge, but do not know whether it is complete. (It is possible that there are some impossibilities we don't know to be impossibilities.)

(P2) If (P1) is true, we are not justified in deeming something possible just *because* we don't know it to be impossible.

(K1) Therefore we aren't justified in deeming anything possible (which we don't know to be true).

(K2) Therefore, talk of possibilities in general is nonsense.

There are additional premises to be added if this argument is supposed to turn out valid. But whatever they might be, I suppose either one of them or (P2) must be false. To see this, compare this line of reasoning concerning our knowledge of

possibilia with our knowledge of *actualia*: on a fallibilistic account (however subtle it is) there is only negative knowledge. Are we ever (in Inwagen's sense) justified in deeming 'All ravens are black' true? Does this affect the fallibilist's notion of objective truth? Is a Critical Rationalist ever justified in believing anything positively? I guess so. But then the case for *possibilia* is as good or bad as for *actualia*, at least on the grounds of Inwagen's argument. It would look like this:

(P1) It is possible that there are some falsities we don't know to be falsities.

(P2) If (P1) is true, we aren't justified in deeming something true just *because* we don't know that it is false.

(K1) Therefore we aren't justified in believing any positive *actualia* (e.g., we aren't justified in believing 'All ravens are black')[26].

(K2) Therefore, talk of positive *actualia* (truth in general) is nonsense.

In Inwagen's example, nothing *excludes* the possibility of one's being able to live on those islands. This latter claim is relative to our currently available information. Since informationalism tries to explain changes in information, it allows for changes in the set of possibilities. Inquiry will bring us "closer to the truth", just as well as it does in a realist's epistemology about actualities. At no particular epistemic juncture are we entitled to claim certainty of our beliefs. Even an *ideal* epistemic situation could not guarantee that our beliefs were not completely mistaken. The same holds for the case of *possibilia*: new information puts constraints on what we deem possible, and inquiry could converge, just as in the case of *actualia* there doesn't have to be a final destination where we would reach certainty. Inwagen's argument again does not show that there is a specific problem with modal knowledge.

4. MODAL ERRORS

We can now take all of Inwagen's arguments in support of modal skepticism to be rejected. Hence, we should try doing some constructive work and see how our proposed explication of modal inquiry given above can yield a classification of errors which, of course, can occur and which should be detectable and correctable if we hope to increase our modal knowledge. Consider the Hesperus-Phosphorus example. Before the information was available that these heavenly bodies are one and the same object (which, it is often said, was a belief held by the ancient Babylonians), it was imaginable that Hesperus would outlast Phosphorus and vice versa. Something we now would describe as *metaphysically impossible*. That it is now known to be metaphysically impossible is the case because we now correlate this statement to another set of information about the relevant issues. There was a time when the identity of Hesperus and Phosphorus wasn't known, at which it was deemed possible that the brightest heavenly body

in the morning sky would outlast the brightest heavenly body in the evening sky. Since we now have information about this issue different from the ancients, it is no longer an epistemic possibility for us. Barwise describes this "modal error" as an error stemming from ignorance.

Another source of modal errors stems from what one counts as relevant in a certain context. The existence of different notions of possibility suggests that for different problems, different issues are relevant – and accordingly, different information is "made" available for the very context in which one claims something to be possible. Hence, another modal error stems from the fact that one can mistakenly overlook available relevant information if one does not realize that an issue is indeed relevant for this very context. I shall discuss one such example in the next part of this paper.

> *The instinctive is just as fallible as the distinctly conscious.*
> *Its only value is in provinces with which we are very familiar.*
> Ernst Mach

> *Intuition is an extremely frail reed upon which to build a philosophy.*
> Robert Nozick

5. CRITICISM AND CONCLUSION

We will now see how our proposed account would model a sample thought experiment and the criticism raised against it. I shall also emphasize what the obvious shortcomings of the proposed account are, and suggest what else would have to be done to provide a satisfying methodology for thought experimentation.

After introducing necessity refuters, I gave an example for a modal argument from the physicalism debate.[27] I said above that physicalism would be rejected by the thought experiment mentioned, if the metaphysical possibility of a zombie world could be accommodated. Various authors seem to hold that the task involves entertaining as possible only a world in which all *qualia* are absent. Sören Häggqvist's treatment of thought experiments, for example, seems to support this view.[28] But according to Häggqvist, if we want to establish a possibility premise for such a modal argument, we stipulate the scenario S and accommodate it in a consistent environment C, where C should be as conservative (as close to the actual world) as possible. This last requirement resembles Inwagen's claim that our modal beliefs are of significance only if bound to our everyday life. In fact, a lot of philosophers hold that the crucial difference between good and bad thought experiments is to be found in their closeness or remoteness from the actual world. I think this misses the point entirely; such a requirement is too inclusive and too exclusive at the same time.

It is too strong since there are mathematical possibilities which I can use in a philosophical or mathematical argument to reject an apparent mathematical necessity, one that can be sufficiently accommodated in a consistent environment that has *nothing* to do with the actual world.[29] What is crucial is their consistency with mathematical truths. A closeness requirement is, on the other hand, too weak, since however far we move from the actual world, it is not the degree of difference that matters, but whether the difference involves overlooking relevant available information.[30] It is for this reason that Barwise and Perry reject the zombie argument against physicalism.[31] Imagining a world that lacks qualia means designing a world in which phenomena are missing that are among the *physical* phenomena of our world (e.g., the physical effects of mental causation). But since at the same time a world is required which is physically *indiscernible* from ours, this bit of information about our world is relevant! Overlooking it means imagining a possible world with no argumentative power against the target claim.

5.1 A Dilemma for Informationalism

Doesn't this seem all too easy? Well, actually it seems to be begging the question. As far as we have studied the informational account, our modal knowledge seems to be *derived knowledge*. We infer the possibilities by constructing models which solve all the issues of certain parts of our background knowledge. Now we can ask whether physicalism itself is part of that background knowledge we take to decide whether zombies are possible. If it is not (and if it is not implied by the relevant background knowledge either), it is trivial to find a model that solves all the issues but is nevertheless incompatible with physicalism (unless physicalism was implied, according to the definition of logical consequence in model-theoretic terms). If, on the other hand, physicalism was implied, it is no wonder we cannot find a possible world contradicting it. Conclusion: thought experiments are either trivial or question begging. Are we back to square one? [32]

5.2 Machianism

One[33] way to escape the dilemma posed in 4.1 is to put evidential weight on *intuitions* as non-inferential beliefs.[34] We would rather leave physicalism out of the relevant background knowledge and activate our intuitions to bridge the gap. Are the (trivially existing) counter-models intuitively satisfying?

Following our epistemological program, we would then have to account for the reliability of intuitions. What are they? Where do they come from? What are the normal conditions for intuiting something reliably? This, I guess, addresses Inwagen's original problem: why should we have a faculty to non-inferentially judge possibilities at all, and why should this faculty be reliable? *Platonism* is

one way to answer all these questions in an incredible way.[35] *Machianism* answers all questions convincingly, but not necessarily satisfyingly. 'Machianism' is a conscious misnomer[36], due to the fact that the better label 'intuitionism' is already in use for another philosophical position. Machianism explains our intuitions as dispositions which evolved during our evolutionary history. Some of them were shaped by the circumstances our ancestors had to struggle with, some were not so shaped. The former can be assumed to be commonly shared within our species. They might include intuitions concerning logical laws[37] as well as fairness intuitions[38] and grammaticality intuitions[39], or intuitions about how to behave in an iterated prisoner's dilemma.[40] When it comes to grammaticality (as in Noam Chomsky's theory), intuitions are all there is to know. If they are commonly shared, evolutionary history explains why this is so, and in which circumstances they are triggered under normal conditions. The biggest plus is that they are reliable in an objective sense: to be commonly shared just is to be true. For all other intuitions this is dubious.

Imagine a moral realist who tests his theory against evolutionarily hard-wired fairness intuitions. Wouldn't these intuitions be a bit too contingent to form the basis for a moral realist's ethics? Is, to take another example, ontology a theory about what there is, or a theory about what evolution made us intuit what there is? As much as this sounds unsatisfying to the realist, as much does it dissolve his object. Robert Nozick, who certainly is a Machian in the sense introduced here, claims the following:

The strength and depth of our intuitions about statements cannot be used as powerful evidence for their necessity when those statements are of a kind such that selection would lead to strong intuitions of their self-evidence even if they were (only) contingently true. [...] If our intuitions of necessity, and our inability to imagine otherwise, can be accounted for without assuming the existence of necessities, should we believe there are any such necessities at all?[41]

Machianism – as we merely sketched it here – explains all epistemological questions concerning intuitions and thought experiments, but it leaves another one untouched: why should we care? Nozick, again, hits the nail on the head:

Philosophers who give great weight to intuitions need to offer some account of why such intuitions are reliable and are to be trusted; at least, they need to sketch how we would have acquired a reliable capacity of this sort. Descartes based his confidence in thought processes that involve "clear and distinct ideas" upon the existence of a good God who would not deceive him. Upon what do contemporary philosophers of intuition base their claims? Of course, if the purpose of such philosophy is merely to codify and systematize the intuitions that (for whatever reason) are held, then a philosophy built upon intuitions will need no further basis. And it will have no further validity.[42]

Is this any better than Modal Skepticism? Wouldn't this lead to the same rejection of modal arguments? I don't think so. Even if one ascribed to Nozickian

Skepticism, thought experiments would still be relevant methodological tools. Let's conclude with what we've found out and see what consequences follow.

We have seen that an important subclass of thought experiments is used as a device to choose between competing philosophical theories. Their methodological place is within theory dynamics. This method crucially involves talk of possibilities and necessities. The *notion* of possibilities and their relevance for theory dynamics is not the methodological problem of thought experiments, contrary to what Peter van Inwagen suggested. Philosophical theories are modal in nature, they quantify possible worlds which *can* count as counterexamples. A broad notion of "possibility" that conforms to our not-impossible-hence-possible inferences is explicable, e.g. in the terms of *informationalism*. "Modal Skepticism" as a local form of skepticism is not supported by any of Inwagen's arguments. The explication of the notion of possibility given here advocates a certain methodological position: what *counts* as a counterexample is not an issue of closeness or remoteness to the actual nature, it is an issue of *relevance*. What is relevant is determined by the modal kind of the target theory itself. Although this seems all too convincing, there still remains the problem of how to deal with philosophical positions (targets and refuters) that cannot simply be inferred from other knowledge we accepted as relevant. So it seems like we need a faculty of non-inferential insight into possibilities. One way to deal with this, *Machianism*, doesn't support rock-bottom relevance for the possibilities we might find intuitively plausible. Nevertheless, it says for some of them what they are, where they come from, and when they are a reliable source of evidence for what.

Imagine a possible world in which philosophers aware of all this would want to keep doing scientifically respectable philosophy. Is there any *special*[43] reason for them to abandon the method of thought experimentation? As the arguments have shown, there are indeed no "special" reasons to give up thought experimentation. So maybe there is a lesson as to how to deal with them? Well, according to our results, there is definitely room to ascribe different roles to the reactions triggered by imaginary scenarios:

(A) Your theory is *about* intuitions with respect to X, not because intuitions are reliable evidence of what X is, but because intuitions *are* the data.[44] You always have to take them seriously, anyhow (if they are in conflict with your theory, you have to find a causal explanation for why this is a deviant case).

(B) Your theory is not about intuitions, but about X.[45] Intuitions conflict with your theory about X. You should try a Machian method of explaining them away. If you are not successful, then that is evidence that these intuitions are on to something.

(C) Your theory is not about intuitions, but about some X that really couldn't interfere with the evolution of intuitions. Additionally, they seem to converge against you. Intuitions are clearly irrelevant, but you might want to find a compromise. This strategy might be necessary if your theory is

ethical in nature and has such counterintuitive consequences that you cannot hope to get them realized in laws (e.g., in a democracy).

NOTES

*   I would like to thank Dieter Birnbacher, Axel Bühler, Jason Hagen, and John Perry for helpful comments on earlier versions of this paper.

1. Like Sören Häggqvist and Peter van Inwagen, I shall refer to philosophers who criticize the method of thought experimentation as 'skeptics'.

2. I cannot argue for that position here, but one can find arguments in Sören Häggqvist, *Thought Experiments in Philosophy*. Stockholm: Acta Universitas Stockholmiensis 1996 and Roy Sorensen, *Thought Experiments*. Oxford: Oxford University Press 1992.

3. Kathy Wilkes, *Real People. Personal Identity without Thought Experiments*. Oxford: Clarendon Press 1988.

4. Daniel C. Dennett, *Consciousness Explained*. London: Penguin 1991.

5. Peter Unger, *Identity, Consciousness & Value*. Oxford: Oxford University Press 1990, Sorensen, *Ibid.*, Häggqvist, *Ibid.*, Tamar Szabó Gendler, *Thought Experiment. On the Power and Limits of Imaginary Cases*. New York: Garland Publishing Inc. 2000.

6. The tradition of this criticism was, of course, present throughout the whole history of philosophy. Nevertheless, it was intensely discussed during the 90's. These methodological worries were probably due to the more and more excessive use of "far fetched" counterexamples in the discussion concerning Personal Identity in the last century.

7. Frank Jackson, *From Metaphysics to Ethics*, Oxford: Oxford University Press 1997.

8. Sorensen, *Ibid.*, p. 19.

9. The term 'thought experiment' is certainly used in a far broader sense in common (scientific) language. I do not wish to construct a taxonomy of thought experiments or to give a conceptual analysis of this term. What I am dealing with is a methodological problem concerning, e.g., this very subclass I characterize.

10. My 'regimentation' differs in several respects from Sorensen's and Häggqvist's. Regimentations are always more or less adequate relative to one's interests. However, the three of us seem to have the same subclass of thought experiments in mind, although we semi-formalize them in different ways. The sentence operators are dummies for 'necessity' ('□') and 'possibility' ('◇'), respectively, in any sense of necessity or possibility whatsoever, as well as in their interpretation in deontic logic or in any other interpretation that is of philosophical interest and validates the argument.

11. I shall say more on this later.

12. Peter van Inwagen, "Modal Epistemology", in: *Philosophical Studies*, 92, 1988, pp. 67-84.

13. Inwagen, *Ibid.*, p. 81: Inwagen claims that there are valid modal arguments that cannot be so described, and refers to the "Gettier cases" – which are, in fact, necessity refuters (Sorensen, *Ibid.*, p. 137).

14. This might be a very complex task to judge sometimes. Coherentism tries to model our reasoning in such situations. See Thomas Bartelborth, *Begründungsstrategien. Ein Weg durch die analytische Erkenntnistheorie*. Berlin: Akademie Verlag 1996.

15. Inwagen, *Ibid.*, p. 71. Note that this denial of the existence of logical *possibilities* is independent of the notions of logical *necessity* or *impossibility*. The latter two are, according to Inwagen, meaningful epistemological notions: "the logically impossible is that which can be seen to be impossible on the basis of logical considerations alone", which is also true of logical necessity. Hence the qualifier 'logical' indicates the source of knowledge for the modal claim. The problem with logical possibility is, again according to Inwagen, on the other hand, that the fact that a certain thing cannot be proved to be impossible by a certain method doesn't make this thing 'possible' "in any sense of 'possible' whatsoever."

16. Another "question like that" would be the question of how we can know that a proposition which we know to be true is also necessary.

17. Stephen Yablo, "Is Conceivability a Guide to Possibility?", in *Philosophical and Phenomenological Research*, 53, 1993, pp. 1-42.

18. At first it seems as if Inwagen wanted to treat basic modal knowledge as non-inferential knowledge (*Ibid.*, p. 70), but Yablo's maxim clearly involves inferential processes. I shall come back to this point in part 5.

19. Inwagen, *Ibid*, p. 78.

20. Is not consistent with (P).

21. Jon Barwise, "Information and Impossibilities", in: *Notre Dame Journal of Formal Logic*, 38, 1997, pp. 488-515.

22. Actually Dretske *defines* information with the use of possibilities. Informationalism doesn't do that. Informationalism explicates the relationship between possibilities and information. Fred Dretske, *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press 1981.

23. Barwise *Ibid.*, p. 491.

24. Barwise *Ibid.*, p. 491.

25. Inwagen *Ibid.*, p. 71-72.

26. Than the ones we know to be necessary on logical considerations alone.

27. Actually, there is more to be said about it, but there is not enough space available to do it here; one might want to turn to John Perry, *Knowledge Possibility and Consciousness*, Cambridge, Mass.: MIT Press 2001 or Daniel Cohnitz, "Two-Dimensionalism and the Metaphysical Possibility of Zombies", in: *Foundations of the Formal Sciences II. Applications of Mathematical Logic in Philosophy and Linguistics*, [Trends in Logic]. Dordrecht: Kluwer Academic Publishers 2002 for more on this problem.

28. Häggqvist, *Ibid.*

29. For examples of the sort of thought experiments I have in mind, one could turn to Lakatos' mathematical counterexamples in Imre Lakatos, *Proofs and Refutations*, Cambridge: Cambridge University Press 1976.

30. This is a point at which we might have a look at physics. Take Einstein's "Clock in the box" thought experiment and his "Travelling at the Speed of Light". The "Clock in the box" was not a successful thought experiment, whereas the latter was. How could you describe the difference between them in terms of "closeness"?

31. Perry *Ibid.* argues in a way different from Barwise *Ibid.*, but the following summary is neutral with respect to that difference.

32. At this point, the difference between Barwise and Perry matters. Barwise's argument really seems to be question-begging, whereas Perry's is actually not. For more on this see Perry *Ibid.* and Cohnitz *Ibid.*

33. Another way to escape from this dilemma is to develop a semantical theory that connects only a fragment of our knowledge (our *pure* semantical knowledge) with relevant possibilities. This is what "Modal Rationalism" tries to do. With the help of this bridge, we would then be in a position to inquire the realm of the relevantly possible without considering e.g. physicalism at issue. Positions of this type were recently defended by David Chalmers, Frank Jackson, and Christopher Peacock. The idea is, roughly, that we have a priori access to certain semantical properties of the statements in question. The so called "A-Intension" or "Primary-Intension" is knowable a priori based just on the fact that we understand the statement in question. Some of these a priori accessible properties (e.g., 'this statement has a necessary primary intension') are supposed to imply relevant possibilities if certain conditions are met (e.g., 'If this statement has a necessary primary intension, then it is either not a posteriori or not metaphysically necessary.'). But it is highly questionable whether this account can be spelled out in sufficient detail. One worry is that the a priori accessible, purely semantical properties are not sharply defined in these accounts. If there is no clear boundary between the contribution of pure semantical knowledge and the contribution of the context to our evaluation of certain statements, there won't be a very content-heavy epistemological point for modal rationalism. Another worry is that there is good reason to believe that the main claim of modal rationalism is false. There are statements that seem to have the relevant a priori accessible semantic properties, but whose other properties do

not correspondent to to the prediction of the Modal Rationalist's thesis. For more on this see Daniel Cohnitz, *The Science of Fiction*, forthcoming.

34. I owe most insight contained here to Johannes Haag. For technical reasons, I take intuitions to be beliefs. They might not be beliefs as defined by George Bealer, but they will remain non-inferential – which is what matters here.

35. For more on Platonism and thought experiments see James Brown's writings on thought experiments.

36. Roy Sorensen linked the following ideas back to Ernst Mach, who certainly wouldn't have liked this too much. Since I could not think of any better name and since this is the *Vienna Circle Institute Yearbook*, I coined the term 'Machianism' for the following position, although my reading of Mach doesn't always coincide with Sorensen's interpretation.

37. Robert Nozick, *Invariances: The Structure of the Objective World*, Cambridge, Mass.: Harvard University Press 2001.

38. Brian Skyrms, *Evolution of the Social Contract.* Cambridge: Cambridge University Press 1996.

39. Noam Chomsky, *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press 2000.

40. Robert Axelrod, *The Evolution of Cooperation.* New York: Basic 1984.

41. Nozick *Ibid.*, p. 125.

42. Nozick *Ibid.*, p. 125.

43. "Special" would be a reason that goes beyond general skepticism. It should be special insofar as it applies to thought experiments "especially."

44. The question of grammaticality in Chomsky and the fairness intuitions of Skyrms are such cases.

45. Consider this example: your theory is about Personal Identity, and there are intuitions according to which Personal Identity, matters to us in a special, irreducible way. John Perry's theory, like most "reductionist" theories, cannot account for these intuitions, so he has to explain them away through a Machian explanation, giving an evolutionary explanation for why these intuitions evolved. See John Perry, *Identity, Personal Identity and the Self.* Cambridge, Mass.: MIT Press forthcoming.

Philosophy Department
University of Düsseldorf
Germany
cohnitz@phil-fak.uni-duesseldorf.de