Philosophical thought experiments: in or out of the armchair ?

Pascal Engel
Université de Genève

Summary

This paper discusses the claim that philosophical thought experiments involve modal claims of possibility, and that these can be analysed as counterfactual claims along the lines of Williamson's analysis (2007). Although basically right, Williamson's view is compatible with the claim that thought experiments deal with conceptual possibilities.

1. *Introduction: philosophical thought experiments*

It is not easy to say what a thought experiment (TE) in science is and what role it is supposed to play: is it essentially an experiment or essentially an exercice in thought? There are, however, clear examples: Galileo's reasoning about motion, Newton's bucket, Einstein's elevator, Schrödinger's cat, etc. TE seem to have their proper home in philosophy, which is widely held to be a discipline dealing with concepts. But in philosophy the examples are much less clear cut. Paradigmatic thought experiments are Locke's on personal identity, Strawson's auditory world in the second chapter of *Individuals*, Putnam's Twin Earth, Jackson's Mary the color scientist, or Searle's Chinese room. But there are more borderline cases: Theseus' Ship, Molyneux's problem, Descartes piece of wax example in the *Second Meditation*, Gettier cases and all the machinery of examples invented by contemporary analytic philosophers for the discussion of particular issues in ethics, metaphysics, philosophy of mind or epistemology[1]. It is not clear, in such examples, whether these are thought experiments because they invovle the exercise of imagination, or because they involv a certain kind of reasoning. For instance is the state of nature of social contract theories a thought experiment? Are medieval reasonings about God and angels thought experiments? We tend to think that they are because we take these creatures – or at least some of them- as fictitious, but the medieval philosophers did not. It is also an open question whether we should conceive scientific thought experiments as experiments which one *could* have really performed if the conditions had been met, or whether they are meant to be for ever imaginary. In contrast, with the scientific case, we have an idea of what an

[1] For a good survey and discussion, see Sorensen 1992.

experiment is. The problem with *philosophical* thought experiments is that they do not seem to be, even remotely, *experiments* at all. They, however, are supposed to play an important role in philosophical argument and theories. Which one? For many contemporary philosophers, their role is to test, in thought, various *intuitions* which are supposed to confirm or to infirm some philosophical claims. The epistemological problem thus becomes that of the status of these intuitions, and of the kind of information or knowledge – if such there be- they are supposed to give us. In this paper, after having reviewed some main conceptions of philosophical TE, I shall mainly discuss the view, recently advocated by Timothy Williamson (2007), according to which a philosophical TE is just a piece of ordinary counterfactual reasoning.

2. Conceptions of thought experiments

Ernst Mach famously defined *Gedankenexperimente* as the capacity to "imagine mentally the variation of facts". He said that this activity is not only available to the scientist, but also to the philosopher, the novelist, and the engineer. But what is their common form?

As the variety of examples given above suggests, it is not clear that they have a common form. But it is at least plausible to suggest that TE have the following structure: to devise a thought experiment involves the conception of a *possible* situation against which we test our intuitions and from which we reason about an actual case. So there is, at the basis of a TE, a *modal claim* to the effect that such and such a situation is possible. Now possibility can be understood in different senses. We can talk of physical possibility, of metaphysical possibility, of epistemic possibility, or of conceptual possibility. It would be circular or question begging to say that physical TE deal with physical possibilities, metaphysical TE with metaphysical possibilities, conceptual TE with concepts. For the question raised by most TE is precisely the question of what kind of possibility claim is involved. In principle a physical TE is supposed to deal with our capacity to conceive physical possibilities, and we evaluate these in the light of what we already know about the physical world. But many physical thought experiments precisely involve going *beyond* what we already know about physical possibilities or to consider whether the envisaged situation does not contradict our existing knowledge of physical facts. They imply an extension of our conception or imagination to new possibilities, and what is in question is whether the possibilities in question are genuine possibilities. But in order to answer the question whether a given situation is possible or not we have to answer another one: how can we have an access to genuine possibilities ? The latter question is *epistemological:* it asks

what sort of faculty or cognitive capacity gives us access to the possible situation: imagination? conceptual understanding? intuition ? *a priori* reasoning ? empirical reasoning? These are not necessarily the same. So the modal question and the epistemological question are closely associated. For if we answer that the envisaged possibility is just a fiction of our imagination, we reject the TE as a "mere" exercice of imagination, and not as a genuine possibility. This problem is also involved in the question raised by Mach, and answered positively by him, of whether TE are a limiting case of real experiments: if they are such, then the possibilities in question are indeed genuine possibilities. So the main issue about any TE is: does conceivability imply possibility ?[2] Much of the success of the method of thought experiments in any domain – physics, ingeneering, philosophy or literature, to name the domains considered by Mach – depends upon how one answers this question.

Let us try to illustrate the proposal that a TE involves a modal claim and an epistemological claim on some examples. Take first Lucrecius' famous TE in *De natura rerum* which asks us to imagine an archer at the edge of the universe, throwing a spear at it. If the spear bounces back, there is a wall beyond the limit, which it iself in space, hence in a limitless medium, and if it does not there is no boundary; hence in both cases the conclusion is that the universe is infinite. This is a clear case where we try to construct a possible situation, in order to test a certain proposition, and where the result is that the situation is not possible at all. The reasoning has the following structure

(1) It is possible that an archer sends a spear at the edge of the universe [modal claim] ;
(2) if in this situation the spear bounces back there is a wall outside the limit and if it does not there is no limit [consequence of (1)] ;
(1) hence in either case there is no limit of the universe.

Lucretius' TE on this view is a *reductio ad absurdum* reasoning from an hypothesis. A question which is left implicit is whether the hypothesis is a product of our imagination, or or our conceptual thought. In the case envisaged it involves imagination but it is also clearly a TE about our concept of space.

For a second example, take Franck Jackson's contemporary famous thought experiment about Mary the scientist, which is obviously reminiscent of Molyneux's question. Mary is a colour scientist who knows everything about the physics and the neurophysiology of colors, but who, having lived in a white and black room with a black and white television screen for all her life, has never actually seen a

[2] For a recent analysis of this classical problem (Descartes, Hume), see in particular Gendler and Hawthorne 2002

colored object. One day she is brought outside and given a coloured screen. Will she learn anything about colors? The structure of the TE is the following.

(1) It is possible that there is a person like Mary knowing everything physical about colour but having never experienced colour [modal claim]
(2) if Mary could experience colour, she would learn something
(3) hence phenonenal properties are distinct from physical properties

Here like in the previous case we are asked to react to a possible situation by using our intuitions about this case in order to validate a certain kind of reasoning. So the question arises of what is the nature and the source of these intuitions.

If we consider the answers which have been given to the latter question for thought experiments in science, there are four main options:

(i) an empirist account, according to which TE are reasonings or arguments based on empirical premises, and which aim at either derive some contradiction within a theory or at proving some consequences of an hypothesis, or to make an inference to the best explanation. On this view, which has been defended most explicitly by John Norton (2004), there is no difference between a thought experiment and an ordinary reasoning from premisses to conclusion, and one criticises a thought experiment just as one can criticise an argument. As Norton says: « Thought experiments are arguments which posit hypothetical or counterfactual states of affairs and which invoke particulars irrelevant to the generality of the conclusion"

The only thing which the thought experiment adds is a contingent illustration which could had been different. For instance Schrödinger's cat or Einstein's elevator could have been a dog, and the elevator could have been a spaceship. So TE are dispensable.

(ii) a Platonistic account, according to which TE are "telescopes pointed onto the world of abstract ideas" (Brown 1991) which give us a access, though some sort of intuition of essences, to a world of Platonic entities

(iii) an imaginarist account , according to which TE do not really bring knowledge, but aim at producing, through the work of the imagination, heuristical fictions, which extend our representational capacities and measure the limits of our ordinary conceptual scheme (Gendler 1996)

(iv) a Wittgensteinian account according to which thought experiments test only conceptual possibilities and are variations on the extension of our concepts.

Although it is clear that Norton's "deflationary" view according to which there is nothing more to thoughts experiments in science than in ordinary empirical reasoning is less easy to transpose to philosophy, since philosophy is usually

thought as a form of conceptual, and as a non empirical kind of thinking. But not everyone agrees with this, and we can nevertheless find parallel views about the epistemology of philosophical thought experiments and of the kind of modal knowledge that they involve.

3. The status of « intuitions »

The idea that we test philosophical claims and philosophical thought experiments on the basis of our "intuitions" is ubiquitous, but also ambiguous. The closest conception to J.R Brown's view of TE for philosophical intuitions consists in conceiving them as intellectual analogues of sense perception, which are the basis of a kind of *a priori* knowledge. This view, which has been held by Gödel for mathematical knowledge, has, as far as I know, never been defended for philosophical thought expriments. The strongest a priorist view is defended by George Bealer (1996, 2002) who advocates what he calls a "modest rationalism". The view counts as rationalist because it says that our knowledge of necessities and possibilities is based on intuitions which are independent of experience, and distinct from ordinary beliefs. Intuitions, on Bealer's view, are intellectual "seemings" distinct from propositional attitudes such as beliefs: you can believe things that you do not intuit (*e.g* that Rome is the capital of Italy), and you can intuit things that you do not believe (e.g. the axioms of naïve set theory). Bealer, however, does not take intellectual intuitions as infallible. They are, on the contrary fallible, and subject to be proved wrong. In other words they are merely *prima facie* justified, although they have to be "modally reliable", that is stable, in order to count as evidence. This why his rationalism counts as moderate (Bonjour 1998 and Peacocke 2004 hold related but distinct views). So, according to such a conception, our intuitive answers to Lucretius TE count as stable, hence as *prima facie* correct answers to it.

At the opposite end of the spectrum of views on intuitions lies the ultra empiricist thesis defended by Steven Stich (1991) and his associates (Stich and alii 2001) for whom intuitions are just empirical judgements. Stich calls "epistemic romanticism" the view that the intuitions elicited by philosophical thought experiments reflect universal epistemic norms. These are supposed by the "romantic" to be tested against our empirical beliefs and to be constant. What we call "intuitions" are actually empirical beliefs. But these beliefs vary highly from culture toculture, and from one socio-economic group to another. Hence they reflect no universal norms, but only variable representations. In order to show this, Stich uses classical philosophical cases, such as Gettier examples (Gettier 1963), in order to show "experimentally" that the intuitions that they elicit are very different

depending upon whether they come from European/ American or from Asian subjects. Gettier examples are standardly taken to show that knowledge is not justified true belief. The experimental set ups consists in presenting Gettier cases such as the following to both Asian and European subjects:

"Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?" (Stich and alii 2001)

European/American subjects give the usual answer: Bob only *believes* that Jill has an American car; whereas the Asian subjects say that Bob really *knows* the proposition. Stich gives all sorts of examples where the intuitions of Asian and European about knowledge are subject to strong variations, and concludes that the Gettier cases in no way show that knowledge is not justified true belief, since Asian accept that accidentally true beliefs might be knowledge[3]. His results, however have been contested (Sosa 2009, Engel 2006): do they show that intuitions about *knowledge* vary from culture to culture, or do they show that the *social attitudes* and the concepts *associated* to knowledge vary from culture to culture? It is not clear that they show anything about the concept of *knowledge*.

The whole debate – *a priori* intuitions or empirical beliefs – seems moot unless one does not clarify the kind of epistemic access one has to the situations described by thought experiments in philosophy. For this we have to understand the nature of the modal claim involved in philosophical TE. In what does our knowledge of possibility consist?

4. Williamson on philosophical thought experiments and counterfactuals

From our *prima facie* characterisation of thought expriments in secion 2, we can derive a simple answer: thought experiments are based on counterfactual reasoning. They consist in reasoning from a "what if?" supposition or in counterfactual terms. The antecedent of a counterfactual describes a possible situation, the consequent describes a situation which we take to be compatible with the consequent when we evaluate the counterfactual as true: "If one threw a spear at the limit of the universe, such and such will happen", "If Mary experienced colour, she would learn something". Lewis (1975) had suggested that fictional

[3] For similar results about many domains of Asian vs Occidental throught, see Nisbett 2003 , which I have criticised in Engel 2007.

discourse is a species of counterfactual resaoning. Williamson (2004, 2007) has proposed the thesis that there is no more, in our modal knowledge of possible facts, than our capacity to handle and evaluate counterfactuals. He bases his account on the familiar equivalence between modalities and counterfactuals defined by Stalnaker (1968):

(a) 'A is necessary' = 'If A were not the case, A would be the case'

$\Box A = \neg A \,\Box\!\!\rightarrow A$

(b) A is possible' = 'It is not the case that if A were the case, A would not be the case'

$\Diamond A = \neg(A \,\Box\!\!\rightarrow \neg A)$

From these equivalences, according to Williamson, we can argue that there is no more, and no less, in our imagining or conceiving possible cases in thought experiments than in our ordinary reasoning with counterfactuals. Counterfactual reasoning is a quite common form of reasoning, which does not rest upon any factulty of intuition of metaphysical possibilities. Many counterfactual reasonings rely upon simple knowledge of empirical regularities. For instance we say "If the bush had not be there, the stone would have ended on the mountain path", or "If there had been an earthquake here there would have been a tsanami", or "If I had been at the airport 5 minutes earlier I would not have missed my flight". Others involve more work of the imagination are are more difficult to assess: "If Hitler had invaded England in 1940, he would have won the war", "If I had been Audrey Hepburn, I would have been slim". Some do not involve imagination at all and are purely tautological "If we had been one more for dinner, we would have been 13". According to Williamson, just as we use counterfactuals to reason about possible states of affairs, we can use thought experiments to reason about possible state of affairs. But we do not use a special faculty of intuition. Neither do we use a special sense of metaphysical possibility. We use of usual cognitive ressources: our empirical generalisations and inferences. So there is no more to the epistemology of thought experiments than to the epistemology of counterfactuals.

Now, how can we apply this idea to the epistemology of thought experiments? It is not enough to remark that we can rephrase familiar TE in couterfactual terms such as:

(1) If my left brain hemisphere had been transplanted into Righty's brain, he would have the same emotions as me

(2) If Theseus's ship had been replaced plank by plank, it would be the same ship
(3) If my world were purely auditory, I would have to rely on the height of sounds to navigate through space
(4) If Condillac's statue had only touch and smell it could not perceive far distances
etc.

In order to see what kind of work the counterfactual analysis does, we have to look more closely at the structure of one philosophical TE. Williamson takes the example of Gettier cases.

It is not immediately apparent that Gettier cases in contemporary epistemology are cases of philosophical experiments. As we saw above, they start from the idea that a necessary and sufficient condition for knowledge is justied true belief, and give counterexamples to this condition by exploiting the elementary logical point that some logical consequences of falsehood are truths and the elementary epistemological point that deduction is a way to transmit justification from the premises to the conclusion of an argument. A subject has a justified belief in a falsehood Q, deduces from it competently a true belief, P, and believes P on that basis. So the reasoning starts form the Gettier condition : justified true belief is a necessary and sufficient condition on knowledge, which can translate thus:

(1) $\forall\, xp\ (\mathrm{K}(x,p) \equiv \mathrm{JTB}(x,p))$

The objection to (1) is the presentation of a Gettier case: It is possible that one has a justified true belief that P which is not knowledge.

(2) $\Diamond x\, \exists p\ \mathrm{GC}(x,p)$

Now what plays the role in the reasoning of the modal claim that I had identified above is the following possibility: if the Gettier case had occurred, then the subject would have had a justified true belief in *p* without knowing , i.e

(3) $\Diamond x\, \exists p\ \mathrm{GC}(x,p)\ \ \Box\!\!\rightarrow\ \ \forall x\, \forall\, p\ (\mathrm{GC}(x,p)\ \supset\ (\mathrm{JTB}(x,p)\ \&\ \neg\ \mathrm{K}(x,p)))$

(where « $\Box\!\!\rightarrow$" denotes the counterfactual conditional "if it were the case that…"

Hence JTB without K is possible :

(4) $\Diamond xp$ (JTB(x,p) & ¬ K(x,p))

But (4) contradicts (1) (Williamson 2004, 2007: 183-187)

The modality is here essential. For the Gettier case is imaginary and considered as possible. It is not a counterexample to the non-modal claim that in fact every case of knowledge is a case of justified true belief and *vice versa (* Williamson 2007: 185). Gettier can claim that his case is possible, as in (2), not that it is actual. Since the 'possibly' qualification is essential in (2), the 'necessarily' qualification is essential in (1) if the objection is to stand.

So on this analysis a thought experiment such as Gettier's is a modal argument going from a modal premise to a modal conclusion. The role of imagination is in verifying the premises. The major premise (3) is a counterfactual conditional. So there is nothing special about intuitions about cases in standard analytic epistemology. It does not rely on a special kind of faculty of insight, which would be *a priori*. It simply relies on ordinary reasoning about counterfactuals, which involves nothing but empirical knowledge.

5. Replies to some objections to the counterfactuality thesis

Williamson's counterfactuality thesis about TE belongs obviously to the family of deflationary views about TE which is advocated by writers like Norton (2004). When he claims that we do not need any spacial faculty of insight into metaphysical possibilities to assess TE, but only the ordinary cognitive reassources involved in counterfactual reasoning. Neither do they rely on a specific conceptual faculty of intuition which would be specific to philosophy. In this respect they do not rest upon any kind of a priori knowledge. Actually Gettier cases have nothing far fetched, and, as Williamson notes, one can construct real life examples of Gettier cases. For instance I can tell you that I have lectured in Algeria, and believing me, you can infer justifiably that I have lectured in North Africa. That belief is true, but the initial one was false: I have never lectured in Algeria, but only in Tunisia.

There are some *prima facie* objections to Williamson't thesis. In the first place, can the Stalnaker equivalences (a) and (b) above show that when we understand statements about possibility we do not understand anything more than the equivalent counterfactual statements? That the Stalnaker thesis of the reduction of modalities to counterfactuals holds does not imply that the epistemology of modalities reduces to the epistemology of counterfactuals. To take another

example of logical equivalence, the fact that ($\neg$ A & $\neg$ B) is logically equivalent to $\neg$ (A $\vee$ B) does not imply that the epistemology of conjunction is the same as the epistemology of disjunction, not any more than when the grocer tells me that he has neither jams nor quamkats I need undersand that it is not the case that he either has jams or quamkats. We do not translate automatically « necessary » in

Socrates is necessarily human

into a counterfactual:

If Socrates had not been human, 2+2 would have been 5

But the objection misfires. The equivalences (a) and (b) do not tell us that we know modal facts *by* knowing counterfactual conditionals. At best they tell us at it is *a way* of knowing modal facts. So the logical equivalences (a) and (b) do not prove the epistemological equivalence of modal knowledge and counterfactual knowledge. It is only supposed to make plausible the claim that our intuitions of possibility and necessity are closely associated with our intuitions about counterfactuals.

Another, related, objection is that Williamson's analysis cannot show that our modal knowledge of the premises of thought expriments (such as (3) of the Gettier reasoning above) is counterfactual knowledge, since the latter actually presupposes the former. We use our intuitions about necessity to handle counterfactual claims. Indeed this is what David Lewis (1973) analysis of counterfactuals in terms of possible worlds and similarity spheres within possible worlds does. When philosophical TE involve far fetched possibilities, for instance TE about personal identity about science fiction cases of teleportation or or helf brain exchanges, what help can we get from the rephrasing of the TE in counterfactual claims?

This objection, however, rests upon a misunderstanding: the counterfactuality claim is not that counterfactuals are *prior* to necessity and possibility claims, but that they come together. The counterfactuality thesis does not say that we understand more easily certain philosophical thought experiments about remote possibilities when we translate them in counterfactual terms. It says that we do not understand them better *nor worse*. In other words, if a thought experiment is unrealistic, it will remain so.

A third objection to the counterfactuality thesis is that it presupposes that counterfactual conditionals have truth conditions and can be evaluated truth conditionally. But this thesis is rejected by those who adhere to the probabilistic analysis of counterfactual, according to which these have only assertion conditions in terms of the conditional probability of the consequent given the antecedent. But

apart from the fact that this analysis is generally taken to work better for indicative than for subjunctive conditionals, it would not fit the structure of TE reasonings, which involve the truth of propositions about possible cases, not their probability. There is another, subtler, reason why the suggestion would be wrong. Typically if one understands conterfactuals in terms of degrees of subjective probability, or in some analysis in terms of the familiar Ramseyan idea that the antecedent consists in adding a belief to our stock of beliefs and we see whether the consequent accords with it, the kind of possibility that we are dealing with is *epistemic* possibility. But this would have the result of making all thought experiments reasonings about epistemic possibilities only. But some thought expriment – actually many – deal with metaphysical possibilities. To reduce these to epistemic possibilities would have the effects of transforming our modal question into a question about our concepts and about our thoughts and beliefs. A Kantian, or a Wittgensteinian, would probably welcome this idea, which certainly fits some TE (in particular Strawson's), but there is no reason to accept it in the first place. For instance TE about personal identity do not deal only with what are our beliefs about personal identity, but with what is, or is not possible metaphysically. Similarly Lucretius did not consider his TE of the archer as dealing only with our thought. In other words there is no reason to limit the range of the modalities invovled to epistemic modalities. If we want to take into account the whole range of thought experiments we have better not presuppose that they deal only with epistemic possibilities.

A fourth objection concerns the nature of the cognitive ressources which are involved in thought experiments. The counterafactuality thesis says nothing about them. It does not tell us whether our cognitive faculties in evaluating counterfactual claims rely on imagination or on conceptual resssources, or simply on empirical beliefs. For instance one might give an analysis of counterfactual reasoning in terms of a simulationist analysis of imagination (Currie and Ravencroft 2002) or in terms of mental models (Byrne 2005). But the counterfactuality thesis does not pretend to decide which of these views is correct. Since it holds that the cognitive ressources involved in thought experiments are just those of our ordinary thinking about counterfactuals, it presupposes that any correct story about how we understand counterfactuals will be useful for understanding thought experiments.

Finally, one might object that Williamson's analysis of the Gettier example is fairly restricted and might not extend to other philosophical thought experiments. But if one agrees with the analysis given above of the Lucretius and Mary TE, the modal claim is quite pervasive. In particular it is quite important to see that they involve possibilities. TE do not propose empirical generalisations which could be tested and which could receive counterexamples. It is quite important here not to confuse the usual pratice of analytic philosophy of giving counterexamples with the practice of constructing thought experiments.

6. Metaphysical vs conceptual possibilities

A more convincing objection to Williamson's analysis is that it has the consequence that no genuine thought experiment bears upon our *concepts* or about our understanding of concepts. All TE, on Williamson's view, bear on metaphysical possibilities: they are about the world, not about the way we think of it. But this seems very implausible, at least concerning philosophical thought experiments. For many of them can be understood as bearing upon our *concepts*, not upon the world. The common wisdom about, for instance, TE on personal identity is that they test our common concept of personal identity. Strawson's purely auditory world in chapter II of *Individuals* bears clearly on our *concept* of an objective world: his question is whether such a concept would be coherent if our experience were limited to a world of sounds only. Many thought experiments explore the limits of the extension of our concepts, whether ordinary or philosophical, and are devised in order to test their coherence. Indeed one dominant conception of philosophical analysis is that it deals with concepts, not things in themselves. TE in this sense serve as a kind of critical tool. This does not preclude TE from being about things and about possible states of affairs, but most of the time we have access to them only through our concepts. Thought experiments about persons or about freedom could hardly be about persons or freedom in themselves, they are first and foremost about our concepts of person and of freedom, even though they are supposed to show us something about persons and freedom. Many philosophical concepts involve theories, or at least a cluster of other concepts which are associated to them.

Williamson's analysis of the Gettier cases as thought experiments is unorthodox in that most analyses take them to bear upon our concept of knowledge, not about knowledge itself, metaphysically speaking. After all this isn't it what the differences between Occidental and Eastern subjects alleged by Stich and his associates purport to show? They diverge on their concept of knowledge, or if one prefers, on the meaning of "knowledge", since the former take it implicity to mean "justified non accidental true belief", whereas the latter seem ready to accept, at least in some cases that justified accidental true beliefs can be knowledge. According to Stich there is no essence of knowledge which could lie behind our intuitions, and the very fact that they conflict shows that no such essence is to be found. The traditional epistemologist, on the contrary, claims that the Gettier counterexamples show that justified true belief is not a necessary and sufficient condition on our concept of knowledge, hence that we have to look for other conditions.

Williamson strongly disagrees with this. He denies that there are conceptual truths and that philosophy is in charge of articulating them. His analysis of thought experiments is in part motivated by his criticism of the epistemological conception of analyticity and of the classicical view that philosophy is an *a priori* discipline. I cannot deal here with his conception of philosophy (see Engel 2009)

What would an analysis of the Gettier thought experiment along conceptualist lines involve? It would imply that (1) above is a conceptual necessity about our concept of knowledge, and that (2) above is a conceptual possibility. But, objects Williamson here,

« The conclusion would be that it is conceptually possible to have justified true belief without knowledge. That does not refute the hypothesis that knowledge just is justified true belief, of metaphysical necessity, any more than the conceptual possibility of something with atomic number 79 that is not gold refutes the hypothesis that gold just is the element with atomic number 79, of metaphysical necessity. The primary concern of epistemology is with the nature of knowledge, not with the nature of the concept of knowledge. If knowledge was in fact identical with justified true belief, that would be what mattered epistemologically, irrespective of the conceptual possibility of their nonidentity. » (Williamson 2007: 205)

Moreover , Williamson argues, the conceptualist reading would be trivial, for on any reasonable understanding of «conceptually possible » it is conceptually possible that some abnormal instance of the GC is not a case of justifed true belief (ibid: 205). In other words would always cook up an abnormal situation which would show that the concept of knowledge and that of justified true belief come apart.

Williamson's argument is is premissed on the view that knowledge is, like gold, a natural kind[4]. But it is not evident that the concept of knowledge is unified and has a real essence in the sense in which natural kinds are supposed to have a real essence. There are borderline cases of knowledge, such as the one which is described in the situation invented by Lehrer (1990) of Mr Truetemp :

Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system of his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been

[4] This is the thesis defended by Kornblith 2003 . see the exchange between Williamson 2009 and Kronblith 2009

inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is?

A knowledge reliabilist, for whom knowledge is but reliable true belief, will accept Mr Truetempt's case as a genuine case of knowledge, even though the situation is, admittedly, a bit abnormal. But an internalist about knowledge will disagree. But it is hard to see how this abnormal case, which is admittedly conceptually possible would be trivial, for it is precisely meant to elicit in us our intuitions about two different concepts of knowledge, the reliabilist and the externalist one. This is, after all, what TE in philosophy are about. Indeed one can agree with Williamson that if Mr Truetemp's case were impossible, this would indeed show something about the *nature* of knowledge, not merely on our concept of it. But this is precisely what is in question! There is a conflict – and a genuine one, as such TE show – between two concepts of knowledge. The TE is meant to illustrate the conflict.

To return to the Gettier case, we can grant Williamson's claim that Gettier cases bear upon *knowledge*, not our concept of knowledge. But that does not imply that we cannot have diverging conceptions of what knowledge is. Internalist vs externalist notions of knowledge, coherentist vs foundationalist notions, or virtue theoretic vs reliabilist notions are familiar oppositions within the philosophy litterature on knowledge. When philosophers use thought experiments in the course of an argument in favour of one or other of these notions, they do not pretend to elicit the true concept just by relying on the intuitions, be they Asian, European, male, female, catholic or buddhist. They intend to *argue* in favour of one conception or other, and the TE are part of their argument. But they are not *all* their argument. This is where both the rationalist a priorist and the experimental philosopher are wrong. The expect to draw out of thought experiment *the* concept, and if they do not find it they conclude, like the eliminativist experimental philosopher, that there is no concept at all, hence (depending on the issue) no knowledge, no persons, no freedom, etc. But this is wrong. For there are philosophical conflicts, which can only be solved (or perhaps can't) through an argument. Thought experiments are not reports about our "folk" concepts
They can be used in order to do that. But even if they can be so used, it does not imply that it is their only use. They are used in philosophy in the context of offering hypotheses, for the sake of arguments. Thought experiments are artifacts within arguments. TE are only in part the reports of discoveries about our beliefs or about our concepts. They are constructed for the sake of an argument. They are hypotheses which we put forward to see what follows. In some cases the

deliverances are convincing. In some other cases not, because there is a conflict between various concepts. The conflict may, or may not exist in ordinary throught.

Discussiong the psychological experiments adduced by Stich and others, Williamson says

«Native English speakers sometimes dispute the Gettier verdict . . . In doing so, they show poor epistemological judgment but not linguistic incompetence: they are not usually accused of failing to understand the relevant words of English; it would be inappropriate to send them off to language school for retraining. »(Williamson 2007: 188)

I agree. But that does not show that a well devised thought experiment, constructed in the course of an argument about knowledge, can elecit diverging conceptions. After all, Putnam's Twin Earth case elicts diverging conceptions of the nature of mental content or of semantic reference. If we had resoleved the issue in favour of externalism contra internalism, or in favour of one conception or other of knowledge, we could certainly say that the Gettier case, or Mr Truetemp case, have offered genuine counterpossibilities, or not. But unless these conflicts are resolved, the question is open. In most philosophical cases it is what happens.

Wouldn't it be nice if philosophy could, like science, reveal essences? Indeed I agree with Williamson that it is what it ought to *aim* at. I do not accept the Wittengensteinian view[5] that philosophy is *entirely* conceptual. But it has to, when a question is not solved. And looking at thought experiments can help in setting the problem,sometimes in solving it. So it does not seem to me that Williamson has shown that all thought experiments deal with metaphysical , and not conceptual possibilities, although I agree that in philosophy we have both.

7. Conclusion

So the scope of the counterfactuality thesis seems to be pretty limited. It actually does not tell us how we can in general assess a thought experiment, or evaluate the claims that it makes. it does not give us a way to decide which of the main conceptions (i)-(v) above of thought experiments are correct. But this is hardly surprising. If the counterfactuality thesis is correct, there are as many kinds of thought experiments that there are kinds of counterfactual reasonings. The counterfactuality claim is not the thesis that we can illuminate the nature of specific thought experiments by translating them into counterfactual claims. It is just the thesis that philosophical experiments do no use ressources beyond those of

[5] Reinstate angrily by Hacker 2009 against Williamson

counterfactual reasoning. And given that counterfactual reasoning is a species of common sense reasoning, this rules out two antogonist views of TE: the ultra rationalist one which says that they appeal to a power of a priori intuition of essence, and the ultra empirist one which says that they appeal to empirical beliefs only. If this is correct, philosophical thinking, in so far as it uses characteristically thought experiments is neither *a priori* nor *a posteriori*. It can be both: sometimes conceptual, sometimes not. [6]

References

Bealer, G 2002 "Modal epistemology and the Rationalitst Renaissance", in Gendler T. and Hawthorne 2002

Brown, J. R. 1991 *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* London : Routledge

Byrne , R. 2005 *The rational imagination,* Cambridge Mass, Cambridge University Press

Currie, G. and Ravencroft I. 2002, *Recreative Minds*, Oxford: Oxford University Press

Engel, P. 2006 « Des avantages et des inconvénients de faire de la philosophie analytique en fauteuil », in M. Ouelbani, ed. *La philosophie analytique*, Université de Tunis

2007 " Is there a "Geography of thought?", *Cognitio*, *Revista de Filosofia*, 8, 2, 197-212

2009 "The Philosophy of the Philosophy of Philosophy", *Logique et Analyse,* to appear

Gendler, T. 1996 *Imaginary exceptions : on the power and limits of thought experiments*, dissertation, Harvard, NY, Garland Press , 2000

Gendler T. and Hawthorne, J. 2002 eds *Conceivability and possibility* , Oxford, Oxford University Press

Gettier, E. 1963 « Is Justified True Belief knowledge ? » *Analysis*

Hacker, P. 2009 "A Philosopher of Philosophy", *Philosophical Quarterly,* Volume

[6] Versions of this article have been read in the Athens conference on Thought Experiments in April 2007, and in the Grenoble conference in June 2008, which was its sequel. I thank very much Sophie Roux and Katerina Hierodiakonou for their invitation, their discussions and their patience, Jean Yves Goffi for his discussion of this article in Grenoble, Stelios Virvidakis, Christophe Grellard, Carla Rita Palmerino and all the participants in these conferences for their remarks. In January 2009 my discussion of Williamson's book at the VAT conference in Tilburg helped me to revise in part my initial views. I thank Filip Buekens, Igor Douven, and the participants for the invitation and the feed back. And thanks to Timothy Williamson for having clarified for me some points discussed in paragraph 5 above.

59 Issue 235, Pages 337 - 348
Kornblith H. “Williamson’s The philosophy of Philosophy”,*Analysis*, 69 1, -109- 116
Lehrer, K 1990 *Theory of Knowledge*, Boulder, Co: Westview
Lewis, D. 1973 *Counterfactuals*, Blackwell, Oxford
1975 “Truth in Fiction”, in *Philosophical Papers*, Oxford: Oxford University Press
Norton, J. 2004 “Why thought experiments transcend empiricism”, in C.Hitchcock, ed. *Contemporary debates in the philosophy of science*, Oxford, Blacwell
Sorensen, R. 1992 Thought Experiments, Oxford, Oxford University Press
Sosa, E. “In defense of intuitions in philosophy”, to appear in *Stich and his critics*
Stich S. 1990 *The Fragmentation of Reason* , Cambridge Mass, MIT Press
et alii 2001 “Metaskepticism” in S. Luper, ed., The Skeptics (Aldershot, England: Ashgate Publishing) 2003, pp. 227
Williamson, T. 2004 “Armchair philosophy, metaphysical modality and counterfactual Thinking” , *Proceedings of the Aristotelian Society* 105, 1, 1-23 2007 *The Philosophy of Philosophy*, Blackwell: Oxford
Williamson 2007 *The Philosophy of the Philosophy of Philosophy*, Oxford: Blackwell