On the Very Idea of a Thought Experiment

Jean-Yves Goffi, Sophie Roux

On the Very Idea of a Thought Experiment[1]

Jean-Yves Goffi & Sophie Roux

Even after the characterisation of thought experiments suggested in the introduction of this book, their population — we are a bit tempted to speak of a bestiary or zoo full of monsters — remains too great for us not to wish to tidy it up a bit: in other words, to distinguish among different types of thought experiment. Several attempts have been made to do so, and we dispose of various taxonomies. You can distinguish thought experiments by the disciplines to which the matter in question belongs, or by their function in a theory, descriptive or evaluative thought experiments, destructive and constructive thought experiments, thought experiments that result in decidable cases and thought experiments that do not, thought experiments that correspond to a physically possible experiment and thought experiments that correspond only to a metaphysical possibility, thought experiments that justify a law and thought experiments that clarify a concept, and so on.[2] Nonetheless, all these taxonomies miss what will prove to be not only a useful distinction to tidy up our zoo, but a response to the radical objection that could be applied to the very idea of a thought experiment. If imagination is free to wander about in a thought experiment, without anything equivalent to the principle of reality able to intervene, is it ever possible that a thought experiment would not work *per se*? Of these two options, only one is real:

— Either all thought experiments are successful, and there is no need to be a stubborn falsificationist to conclude that the very idea of a thought experiment should be abandoned.

— Or there is a distinction between successful and unsuccessful thought experiments, and the idea of a thought experiment could have a certain epistemological depth.

---

[2] For examples of taxonomies, see Popper, *Logic,* 442-452; Brown, *Laboratory*, 33-48; Sorensen, *Thought Experiments*, 197-201; Rescher, *What if,* 47.

In that sense, showing that there is a distinction to be made between successful thought experiments and unsuccessful thought experiments and explaining why there is such a distinction aims not only at making a distinction for the sake of 'taxonomophilia'. The point is also to show that the idea of a thought experiment is a substantive idea that has a certain epistemological depth.

Distinctions between legitimate and illegitimate uses of thought experiments could already be encountered in earlier philosophy of science.[3] But our question about the conditions for successful thought experiments has rarely been addressed in recent literature, in particular if it is seen as a way of defending the very idea of thought experiments. To answer this question, we will proceed in three steps. First, we will expose the problem and present our general method to answer it. Secondly, we will analyse some examples of famous successful thought experiments in order to formulate a hypothesis on the conditions of a successful thought experiment. Thirdly, by offering examples of unsuccessful thought experiments we will illustrate what happens when one of these conditions is not fulfilled. In conclusion, we will make a few remarks on the whys and wherefores of our argumentation.

1. The problem and the method

Let us be granted our impression that there are thought experiments that work. Inasmuch as thought experiments are arguments, "to work" here means, at least to begin with, nothing other than to be successful as an argument, in other words to be able to convince one´s interlocutors. A first caveat is necessary here: writing that thought experiments are arguments might appear as an unjustified bias against Tamar Szabó Gendler, who recently argued against John Norton that at least some thought experiments are not justifiable on the basis of straightforward arguments.[4] It might be an unjustified bias, but in this paper we do not need the strong Nortonian thesis that thought experiments are *only* arguments: we will be satisfied with the thesis that they are arguments, a thesis that even Gendler would not refuse. Indeed our question is simply the following: inasmuch as

---

[3] For example Duhem, *La Théorie,* 306-309; Popper, *Logic,* 442-456.
[4] Gendler, "Galileo"; Gendler, "Thought Experiments Rethought".

thought experiments are arguments, and notwithstanding the fact that some of them might involve a contemplation of an imaginary scenario, how is it that some of them work, while others do not?

In one of his papers, Norton asks precisely the question about the mark that could identify successful thought experiments; he ends up writing that "the mark is just that the thought experiment either uses an argument form licensed by a logic or can be reconstructed as one".[5] But, since every argument must respect standard logic, this cannot be a sufficient mark for identifying successful thought experiments: they have to satisfy more precise conditions. Nor would a reference to truth help here: it may happen that shared assumptions between the interlocutors yield false conclusions. Note in particular that a thought experiment can be successful in the sense that a majority of people admit that it has a certain outcome that is considered as intuitively obvious, although later experiences or further theorising will prove that it is not correct: this is for example the case of the probability paradoxes, well known to every college freshman. In the conclusion of this paper, we will come back to the issue of relativism that our perspective may raise; to begin with, we will, however, not consider this issue, but focus instead on the question of how one can be convinced by some thought experiments and not by some others.

We find in the secondary literature various hypotheses to account for the feeling that there is a difference between thought experiments that are convincing while others are not. According to a first hypothesis, in the case of thought experiments that work, laws of nature or even natural kinds are preserved.[6] According to a second hypothesis, our intuitions ground thought experiments that work.[7]

However, we feel that the question as to whether laws of nature are preserved is not pertinent. If the thought experiment in question takes place in the realm of the natural

---

[5] Norton, "Why Thought Experiments," 53-54. Our departure from the question of the nature of thought experiments to the question of the conditions they respect when they are successful is also to be found in Janis, "Thought Experiments Fail," 113-118; as for Atkinson and Peijnenburg, "Thought Experiments Poor," they show that unsuccessful thought experiments, whether in science or in philosophy, share the two same features: they lead to contradictory conclusions or to conclusions that beg the question.

[6] This is a point made by Wilkes, *Real People*, 12-15, *passim*, against thought experiments concerning personal identity, most notably Parfit's people splitting like amoebas. See also Jamieson, "Method," 484-486; Brown, *Laboratory*, 30-31.

[7] Brown, *Laboratory*; Gendler, "Galileo," and Gendler, "Thought Experiments Rethought," albeit with completely different conceptions of intuition.

sciences, then it would be, at the least, unsatisfactory to speak of preserving the laws of nature without further specification, for what one is precisely seeking to establish by means of this thought experiment might be itself a law of nature. If the thought experiment in question does not concern the natural sciences, then the laws of nature could be preserved, and yet no intuition may be at hand, because what is at stake is not a law of nature.

The second hypothesis — that thought experiments that do not work leave us lost in terms of intuition — does indeed correspond to the feeling one can experience faced with certain thought experiments. We would like, however, to know a bit more about this feeling and what can be understood by 'intuition'. Certainly, speaking about intuition does not necessarily imply that we attribute to people a faculty of direct insight into supersensory truths: experience of things and people indeed yield some intuitions as to what two equally heavy bodies are, how most people would behave in a given situation, when a body is said to be faster than another, what it is to be a good person or to be a bad person, etc. But, even granting this, intuition is a notoriously ambiguous notion. Moreover, it would remain to determine what an intuition that intervenes in a thought experiment is, and how it could be not only psychologically persuasive, but convincing in a controversial context.

This is what we will describe in this paper, using a method inspired by a counterfactual thought experiment presented by Nicholas Rescher.[8] We use his example because of its great pedagogical value, but it will soon be clear that its strong counterfactual dimension is not what is most significant about it. Let us begin with the principle that a counterfactual thought experiment is in reality counterfactual reasoning — to be distinguished from counterfactual declarations, which pose totally different problems and raise totally different questions. In counterfactual reasoning, one introduces a supposition held to be false: "Caesar never crossed the Rubicon"; "General De Gaulle was killed in 1940 during the battle of France". This is the beginning of a "what if" scenario. What follows is the development of the consequences of this supposition: "the Roman Republic is not overthrown", "the Resistance is dominated by the Communists". Rescher wondered just what happens in these cases. The principle of his response is as

---

[8] Rescher, *What if,* 98-99.

follows: the counterfactual arguments aim to determine the consequences of the introduction of a supposition that contradicts our beliefs, that is to say, that is in conflict with all we assume to be true, at least in one region of our knowledge, a region as broad or narrow as we wish. The difficulty is that, if these initial beliefs are consistent, we cannot preserve them all under the supposition that we have replaced one of them by a new belief that is the negation of the first. The issue is thus, in Rescher's words, to "break the chain of inconsistency at its weakest link". Let us reason, following Rescher, about a well-known example and suppose that we introduce the following counterfactual: "If this rubber band were made of copper…".

Well, if this rubber band were made of copper, what then? There are two possibilities, the first one 'natural':

(a) If this rubber band were made of copper, then it would conduct electricity (since copper conducts electricity);

the second one 'unnatural':

(b) If this rubber band were made of copper, then it would not conduct electricity

(since this band does not conduct electricity).

Why do we consider (a) a more natural conclusion? We can explain this by making explicit part of the beliefs of the 'experimenter'.

(1) This band is made of rubber.

(2) This band is not made of copper.

(3) This band does not conduct electricity.

(4) Things made of rubber do not conduct electricity.

(5) Things made of copper do conduct electricity.

(1)-(5) are a consistent set of beliefs in the textbook sense: it is possible for all the statements expressing them to be true. If we introduce as a counterfactual supposition the negation of (2), i.e. non-(2), we must immediately reject (2) by definition, as well as (1) – for the reason that the same object cannot be both copper and rubber (at the same moment and in the same proportions). Non-(2) was introduced at the price of evicting (1) and (2): what else can be kept or rejected? We can keep (4) which expresses a general law relative to the conductivity of rubber, and whose value of truth is not affected by the choice of non-(2) rather than (2). We can still choose to keep (3) since it states a true fact relative to

this rubber band (it does not conduct electricity). But we would then have to reject (5) because non-(2), (3) and (5) cannot all be true at the same time. The choice of keeping (3) thus leads us to the 'non-natural' conclusion. The 'natural' conclusion will therefore be interpreted in the following fashion: by introducing non-(2) we reject (2) and (1). We can keep (4) for the same reasons as before. And this time, we choose to keep (5), since it states a true fact relative to things made of copper. But if we decide to keep (5), we have to eliminate (3) because non-(2), (3) and (5) cannot all be true at the same time.

We have employed expressions such as "we decide to keep" or "we choose to preserve". This gives the impression that the process, at a critical point, is fully arbitrary. The issue is not with the initial supposition: that is part of the normal regime for thought experiments, because of the counterfactuality clause. No, the problem lies rather in the way the procedure plays out to determine the weakest link; why reject (3) rather than (5)? Rescher has a general answer to this objection: it's a matter of cost-benefit optimisation relative to one's overall commitments. In the case of the rubber band made of copper, the optimisation is obvious: Rescher remarks that it would cost more to dismiss general statements about copper things like (4) and (5) than particular statements such as (1), (2) and (3). It is therefore not arbitrary to seek to preserve the former, and to find the weak link among the latter.

The lesson we intend to learn from this example is of general order. We will summarise it in three points:

i)   Roy A. Sorensen and Soren Häggqvist resort to modal logic to differentiate different types of thought experiments.[9] It is indeed legitimate, given the equivalence of counterfactuals and modalities established by Stalnaker.[10] But we are not interested in this; rather, we aim at understanding what pragmatic conditions successful thought experiments satisfy. For that, modal logic is not necessary, since we succeed in dispensing with it.

ii)   Thought experiments are procedures that resemble simple tests of consistency. However, they not only reveal that the whole of the experimenter's beliefs have become inconsistent, they also orient him

---

[9] Sorensen, *Thought Experiments,* 133-166, and Häggqvist, *Thought Experiments,* 92 sqq. For a shorter presentation of Sorensen's theses, see Sorensen, "Thought Experiments and Epistemology".

towards an alternative in which each of the branches has an acceptance/rejection basket.

iii) Thought experiments invite the experimenter to seek the weakest link in her body of beliefs. It is in the ordering of the beliefs, not necessarily perceived before the thought experiment was carried out, that the thought experiment will make the weakest link obvious.

In this sense, a thought experiment is a test whose goal it is to reconfigure the organisation of our beliefs. In that, it is no different from a real experiment. But the confrontation with physical reality is not necessary: the reconfiguration can happen even following a counterfactual scenario.

2. Successful thought experiments

Equipped with the basics provided by Rescher's example, we can enter the zoo of thought experiments and try to understand the reasons some of them are successful. We will now visit an extremely familiar animal, a Galilean pet; we will then climb on the shoulders of Ernst Mach, albeit to advance theses distinct from his; lastly, we will seek confirmation of our theses by heading to the other side of the zoo to visit an unconscious violinist.

A familiar animal: Galileo's two bodies strapped together[11]

The domain of this thought experiment is physics. It not only reveals the inconsistency of a certain set of beliefs, it orients the experimenter towards a new belief. This is why it has been described as being simultaneously destructive and constructive:

— destructive aspect: to refute the Aristotelian proposition (a)

(a): two bodies of the same material descend in the same medium with speeds proportional to their weights;

---

[10] For a short reminder on this question, see Engel's essay in this volume.

[11] This thought experiment was already presented in the *De motu*, but we will refer here to the *Discorsi*, in Galileo, *New Sciences*, 107-109. As Palmerino 's essay in this volume reminds us, it is probably the most widely commented thought experiment in literature. See for example Popper, *Logic,* 442-443; Koyré, "*De Motu gravium"*, 224-232; Brown, *Laboratory*, 77-79; Brown, "Thought Experiments," 122-126; Sorensen, *Thought Experiments*, 126-127; Norton, "Are Thought Experiments," 340-345; Gendler, "Galileo," 397-424; Atkinson "Experiments"; Kühne, *Gedankenexperiment,* 31-57, *passim*.

— constructive aspect: to replace it by the Galilean proposition (b)

(b): two bodies of the same material descend in the same medium with like speeds.[12]

What happens if, in a world where (a) is a law of nature, we take a heavy body A and a light body B, strap them together and drop them from a certain height? Our presentation of the thought experiment follows step by step the dialogue between Salviati and Simplicio. Note that the bone of contention between recent commentators does not concern its logical reconstruction, which is fairly consensual, but the nature and function of the assumptions necessary to draw a conclusion.

(1), from (a): the heavy body A and the light body B fall with speeds $v_B < v_A$;

Simplicio accepts the assumption (A_1): natural speed is mediative, i.e. that if a slower body is connected to a faster body, then the slower will slow down the faster and the faster will speed the slower; hence (2), from (1) and (A_1): $v_{A+B} < v_A$.

But a broad and sweeping assumption, here formulated, is (A_2): weight is additive, i.e. that the weight of a body combined of two bodies is the sum of the weights of these two bodies, $W_{A+B} = W_A+W_B$; hence (3), from (A_2): $W_{A+B} > W_A$; and finally (4), from (3) and (a): $v_{A+B} > v_A$.

From the contradiction of (2) and (4), we infer non-(a): two bodies of the same material do not descend with speeds proportional to their weights. We consequently give up (a) and substitute for it (b): two bodies of the same material descend in the same medium with like speeds. In order to make this last step, we need a last assumption (A_3), that weight is the only parameter to take into account when we study falling bodies; the warrant for (A_3) is the fact that the initial Aristotelian proposition (a) concerns bodies from the same material falling in the same medium.[13]

From a logical point of view, this thought experiment is a *reductio ad absurdum* of (a). This *reductio ad absurdum* works if, and only if, we admit assumptions (A_1), (A_2)

---

[12] Brown, *Laboratory*, 77-79, and Brown, "Thought Experiments," 124, called "Platonic" the thought experiments that are simultaneously constructive and destructive; in so far as they are constructive, they would yield an *a priori* knowledge of nature.

and (A_3), which is not obvious from a physical point of view, the existence of alternative physical scenarios being enough to destroy the Brownian view that this would be a 'Platonic' thought experiment that gives us the occasion to acquire an *a priori* knowledge of laws of nature. For example, against assumption (A_1), that natural speed is mediative, we could say that (a) concerns only *natural* bodies that have *natural* speeds[14]. But strapping (or tying or gluing) two bodies together does not make one *natural* body and that, consequently, (A_1) simply does not apply — we may know a law concerning the fall of natural bodies, not a law concerning the fall of artificially attached bodies[15]. Or, against assumption (A_3), that weight is the parameter to take into account in the case of falling bodies, we could say that we have to take into account the configuration of bodies, as well (which is trivially true: a ball and a plank of wood do not fall the same way) — or variations of the gravitational field (which is true if you are a 20th century physicist) — or colours and odours of bodies (which is not true according to Aristotelian, Galilean or Einsteinian physics).[16]

Our point here is of course not to charge Galileo with logical weakness or with physical incompetence. Nor is it to determine if an Aristotelian of the 17th century would have shared the view Galileo ascribes to his Aristotelian Simplicio. Rather, it is to ask the question as to why the Galilean Salviati can plausibly present this argument as a "short and conclusive argument" "even without further experiment" to the Aristotelian Simplicio.[17] According to us, if Salviati can present his argument as conclusive, it is because:

> i) (A_1) (A_2) (A_3) are assumptions that both the Aristotelian and the Galilean of the dialogue agreed upon.

[13] Galileo, *New Sciences*, 105-106. Brown *Laboratory*, 78, and Brown, "Thought Experiments," 125, claim that the Galilean theory is not logically deduced from known data precisely because he does not take this premise into account.

[14] Galileo, *New Sciences*, 105-106, makes clear that speed is determined by nature.

[15] This objection is raised by Koyré, "*De Motu gravium*," 231-232. See as well Gendler, "Galileo," 405-406, who objects moreover that an Aristotelian would introduce a distinction between *united* and *unified* bodies, or say that the speed of fall depends on the degree of connectedness between the strapped bodies.

[16] Atkinson, "Experiments," 221-222 considers a world in which there are variations of the gravitational field; Atkinson and Peijnenburg, "Galileo," 128-133 examines moreover situations of falling in resisting or turbulent media.

[17] Galileo, *New Sciences*, 107.

ii) Both of them consider that it would cost more to give up one of these basic assumptions than to give up (a).

Can we go farther and explain why Simplicio and Salviati agree on certain assumptions and judge that these assumptions are stronger propositions than (a)? In the case of Rescher's rubber band, the priority between beliefs was easy to determine because it amounts to the priority of one general statement (copper conducts electricity) over a particular statement (this band does not conduct electricity). In the case of Galileo's strapped bodies, the process of prioritisation is not that obvious: for example, (A_1) could have been taken as a premise in the construction of a concept of speed.

Some zookeepers: Mach and beyond

To the question we just asked, Mach gave at first sight an interesting answer, namely, that the two protagonists of the *Discorsi* have stores of instinctive knowledge that comes from our former experiences with bodies. In his masterful work, *The Science of Mechanics*, Mach asked an important question: how is it that the result of a thought experiment may have more authority than the result of a real experiment? The answer that, as a good empiricist, he gave to this question is that a thought experiment taps into a store of instinctive knowledge and summons up beliefs from this store with respect to a specific problem. As explained in the introduction of this book, Mach's notion of a thought experiment is however a complex one, that needs some clarification before being used. This has not prevented the vulgate that claims affiliation with Mach to say that there are intuitions that intervene in thought experiments and to attribute to them in a very general fashion the following characteristics:

i) Not only are they tacit beliefs but, insofar as they result directly from the silent impression produced by nature in the space of our representations, they are barely even propositional attitudes.

ii) Taken together, they constitute a shapeless whole, contrary to the organised structure that is that of constituted science.

iii) They are primarily common intuitions coming directly from our innate knowledge or from basic experiences with bodies.[18]

But, if the intuitions that intervene in thought experiments had these three characteristics, it would be impossible to understand the mechanism of thought experiments. Better to say that:

i) A thought experiment does not have as its material any representation that would be the direct impressions that nature leaves in us, but rather, beliefs. It is usually said that beliefs intervening in thought experiments are 'implicit' or 'tacit', but they are not implicit in the sense that they would constitute a kind of mysterious knowledge. They are implicit simply in the sense that they are unexpressed, while being 'close at hand', to use the Machian expression: they may be there, just under the surface, the function of the thought experiment being to make them come to the fore[19]. For example, in the Galilean thought experiment, assumptions (A_1), (A_2) and (A_3) are of the same nature, even though (A_2) is not expressed: it would have been if, contrary to what our thousand-year-old experiences with balances and other simple devices show, the 17th century Aristotelians had defended the thesis that weights are not additive.

ii) Our store of intuitive knowledge is not shapeless, but has a hierarchical organisation. By this, we mean that there are links between beliefs and that there is a distinction between stronger and weaker beliefs. The evidence for this hierarchical organisation is that when we abandon one belief, other beliefs fall with it, and that we at times prefer to give up certain beliefs so as to be able to keep others. Retrospectively, we say that the beliefs we gave up were 'weaker' than the 'stronger' ones we kept, or, as we did after Rescher, employ the metaphor of cost-benefit optimisation.

---

[18] Gendler, "Galileo," 414-415, and Gendler, "Thought Experiments Rethought," 1161-1162. In the conclusion, we explain more thoroughly the reasons why we cannot agree with Gendler.

[19] John Zeimbekis objected that the relevant belief for a thought experiment I am performing may not be already there in my set of beliefs: I don't after all have infinite beliefs, i.e. I don't have all belief–or disbelief–states corresponding to *all* possible well-formable propositions from the concepts I possess. True enough: we cannot decide if thought experiments help beliefs to crystallise or simply to come to the fore. But we will let the question undecided, because a justified decision would lead to a discussion on tacit knowledge that we cannot afford here.

iii) You must accept all kinds of beliefs in these stores of knowledge, and not only beliefs inspired by a passive observation of nature or by an elementary manipulation of bodies. To confine ourselves to the domain of physics, about which Mach was speaking, the latest results of more formalised sciences may intervene as beliefs just like the know-how of engineers. In other words, the distinction between 'commonsensical' beliefs and 'sophisticated' beliefs is not a relevant distinction here; the only relevant distinction is that between what is accepted as common ground by performers of the thought experiment, and what is not. This is why Galileo makes explicit that the Aristotelian Simplicio agrees to Salviati's assumptions. If the extension to 'sophisticated' beliefs is granted, as we think it should be, it can be understood why there are thought experiments not only at the more primitive stages in the history of mechanics, but at other stages as well.

In this respect, it is interesting to come back to the example that initiated Mach's reflections on thought experiments in the *Science of Mechanics*, Stevin's necklace. Remember that the crucial assumption for this thought experiment is the impossibility of perpetual motion: if this impossibility were not granted, there would be nothing to work on, the other assumption (when one removes equal things from equal things, one has equal things) being untouchable. Not surprisingly, Mach claimed that it is something known instinctively:

> Unquestionably in the assumption from which Stevin starts, that the endless chain does not move, there is contained primarily only a *purely instinctive* cognition. He feels at once, and we with him, that we have never observed anything like a motion of the kind referred to, that a thing of such a character does not exist. This conviction has so much logical cogency that we accept the conclusion drawn from it […] without the thought of an objection, although the law, if presented as the simple result of experiment, or otherwise put, would appear dubious.[20]

Mach may lead the casual readers astray here. The impossibility of mechanical perpetual motion would not be granted by everybody — not only insomuch as there will always be a nut who will believe in something aberrant, but because this principle has a history.

---

[20] Mach, *Science of Mechanics,* 34. Following in Mach's footsteps, Gendler, "Thought Experiments Rethought," 1161-1162, claims that this thought experiment "evokes certain quasi-sensory intuitions" and that the resulting belief is produced "not inferentially, but quasi-observationally".

This history remains to be written, but its main lines are clear: during the late Middle Ages, many projects of perpetual motion were elaborated; during the 16th and 17th centuries, engineers and scientists began to treat the impossibility of mechanical perpetual motion as an assumption needed in physical proofs; this assumption was finally the object of a demonstration in Lazarre Carnot's *Essai sur les machines en général* (1783). The belief that mechanical perpetual motion is impossible, that Mach treated as an immemorial instinctive knowledge common to all men, thus emerged slowly in history and was at the beginning of the 17th century a rather sophisticated belief, admitted only by engineers and scientists.

Hence, the outcome of our jaunt with Mach is twofold. First, the Machian notion of instinctive knowledge is not clear enough to be a real explanation of why certain processes of prioritisation happen rather than others; better to leave it unexplained. Secondly, what Mach called "instinctive knowledge" and that his followers call "intuitions" may be more aptly described as beliefs, in order both to avoid the mysticism surrounding the notion of intuition and to include more or less elaborated kinds of statements. This might be enough to suggest that a thought experiment works not because it operates in one field or another, but rather because is has as its material certain beliefs, beliefs that are shared and that constitute a hierarchically organised whole. Before coming to a fuller presentation of our hypothesis, we will test it by leaving the menagerie of falling bodies of the 17th century to consider a famous 20th century freak, Judith Jarvis Thomson's unconscious violinist.

The exotic aviary: Thomson's Violinist[21]

The domain of this thought experiment is applied ethics. It is an argument against opponents to abortion, the outcome of which is to distinguish between two senses of the expression "right to life".

Thomson begins by a reconstruction of her opponents' argument:

Point (a) grants that the foetus is a person from the moment of conception, *i.e.* she gives her opponents' thesis its greatest strength. Point (b) consists in a kind of syllogism:

---

[21] Thomson, "Defence of Abortion".

every person has a right to life; but a foetus is a person; hence a foetus has a right to life. Point (c) is a bit more difficult. It begins with a set of premises, from which a conclusion is drawn.

Two broad and sweeping assumptions are necessary to begin with:

(A_1) There are differences between the stringency and strength of rights.

(A_2) It is morally wrong to give more weight to a right whose stringency and strength are weak than to a right whose stringency and strength are great.

Thomson's opponents' argument is the following:

(1) A woman has a right to decide what shall happen in and to her body (right to autonomy).

(2) A person's right to life is stronger and more stringent than another person's right to autonomy.

(3) From point (b): A foetus is a person.

(4) An abortion (the fact of voluntarily killing a foetus) performed in the name of autonomy is a case of giving more weight to a right whose stringency and strength are weak than to a right whose stringency and strength are great.

Hence, they conclude:

(C_1) Abortion is morally wrong (Evaluative conclusion).

(C_2) An abortion may not be performed (Practical conclusion; another broad and sweeping assumption is at work here: (A_3) Morally wrong acts may not be performed).

Thomson's strategy consists in performing a thought experiment showing that (2) is the weak link in this set of premises; more precisely, showing that, if there is an equivocation in (2), then the argument is not sound.

Suppose you wake up one morning and find yourself back-to-back in bed with an unconscious violinist. He suffers from a fatal kidney ailment and the Society of Music Lovers has discovered that you alone have the right blood type to help. They have therefore kidnapped you and plugged his circulatory system into yours, so that your

kidneys can be used to extract toxins from his blood as well as yours. But the Director of the Hospital is very comforting: "Put your mind at ease, it's only for nine months".[22]

This is a clear case of treating a person as a means only, not only an intuition that the members of the Society of Music Lovers are behaving strangely or rudely. And it is premise (2) that falls under attack: broad and sweeping assumptions (A_1) and (A_2) are common to both Thomson and her opponents, as is premise (1).[23] Premise (3) has to be agreed on since it is the conclusion of a sound argument. True, premise (4) is not accepted, but only because (2) has been rejected to begin with, and as a consequence of this rejection.

One can rationally refuse to admit (2) because (2) mixes up two senses of "right to life". "The right to life" can mean either "the right not to be arbitrarily killed", or "the right to receive full support to life". Consequently, (2) may mean:

— either: (2^a) a person's right not to be arbitrarily killed is stronger and more stringent than another person's right to autonomy (which seems reasonable enough as it boils down to saying that I may not wantonly kill someone as an expression of my personal autonomy: actually, it is the core idea in the good old Lockean theory of the right to life).

— or: (2^b) A person's right to receive full support to life is stronger and more stringent than another person's right to autonomy (which may be true under certain circumstances, but which is not true without serious qualifications, so that the burden of the proof is still incumbent to the members of the Society of Music Lovers).

A few qualifications may be added to this logical presentation. First, one could wonder how it is possible that an unconscious violinist desperately needs another's body for life support: here, it is the possibility of the scenario – that is to say its credibility – that is in

---

[22] The initial clause, "suppose you wake up one morning and find yourself back to back in bed with an unconscious violinist" is a rhetorical artifice: in order to understand this thought experiment, there is no need for the reader to take on the situation and the reactions of the person who would wake up like this one morning. This differs from the case of Jim and the Indians recalled in the introduction of this book, where the reader must, in a certain fashion, become Jim to understand what is in reality the deep indifference supposed of Utilitarians with respect to the question of personal integrity, and which raises issues for them.
[23] The scenario says: "No doubt, the mother has a right to decide what shall happen in and to her body; everyone would grant that".

question. The response can be relatively direct: an unconscious violinist may desperately need another's body in the sense that the victim of a car accident may have a desperate need for a transfusion with a rare blood type. In this case, urgency may push doctors to very strongly insist that any staff member with this rare blood type accept to donate their blood: Thomson proceeds by passing to the limit of these well-known cases, and this seems a respectable procedure.

More seriously, it may be noted that Thomson's thought experiment relies on an analogy between a foetus and an unconscious violinist. The obvious objection is that this analogy is not obvious: the relation of the foetus to the pregnant woman is not of the same nature as the relation of the comatose violinist to a perfect stranger. But in the same way that in real experiments we do not take into account all that is actually encountered, likewise a thought experiment does not have to take everything into account. At issue is the extension of the concept of "the right to life". The very 'logic' of the language of rights is to clearly distinguish between virtual ethical and legal rights on one hand, and real psychological and affective relations on the other. One can, of course, deplore this. But that does not go against Thomson's thought experiment, since the idiom of rights is hers, but also that of her adversaries.

Lastly, we can wonder what would happen if we were dealing with someone who believed that it is proper, given these imaginary circumstances, to authorise the Society of Music Lovers to make use of the body of another, against his will, for at least nine months, so as to permit the comatose violinist to continue to play, as her exceptional musical talent is a divine gift. There is no point in stating here that such a conviction would be excessive, for the response is only too predictable: 'it is not because an opinion is held by a minority that it is excessive'. But the answer can be as follows: Anyone who would speak in this way would put herself out of bounds, so to speak, with respect to the language of rights. She would adopt another argumentative posture, admitting, for example, that life is a gift of God, and that human beings must act in consequence. But beyond the fact that such an affirmation has its own well-known weaknesses,[24] it is not pertinent with respect to everything that had been admitted by the interlocutors when the agreement on the premises was slowly but surely established.

---

[24] It is sufficient to refer readers to Pabst Battin, *Ethical Issues*.

Thomson's thought experiment reveals that, just as in public or executed experiments, thought experimenters need to share beliefs and to hierarchise beliefs if they want to reach a conclusion: in this respect, a thought experiment may be a powerful tool in showing which beliefs are shared and which are not shared, appearances notwithstanding. More generally, our analysis of two well-known successful thought experiments and our discussion of Mach led us to formulate the hypothesis that a thought experiment works under the following conditions:

(I) At issue in the first condition is simply setting off the domain in which one is situated. A thought experiment deals with beliefs, not with things, nor with these mental representations of things that are supposed to be intuitions.[25] There is nothing to wonder at in this: any experiment whatsoever tests beliefs, the difference between a thought experiment and a physical experiment lying not in what is tested but in the way the test is performed.

(II) A thought experiment deals with beliefs, not with things, which in our examples we called 'assumptions' and which can perhaps more properly, insofar as there are not always explicit assumptions, be designated as the 'context' of the thought experiment[26]. There is nothing to wonder at in this. The existence of shared assumptions or of a background context is a condition for argumentation in general.

(III) These beliefs should be organised as it were in a hierarchical structure, and this hierarchical structure should be acknowledged by interlocutors, so that, in case of an inconsistency, they will agree to judge that it is better to abandon one belief rather than another. It turns out that this is the main condition for a successful thought experiment.

3. Unsuccessful thought experiments

The examination of thought experiments that do not work will now allow us to argue that each of these conditions is individually necessary for a thought experiment to work. We

---

[25] Kuhn, "Function for Thought Experiment," 242, 261, first pointed out that thought experiments test beliefs ('conceptual' or 'mental apparatus' in his own terms) — which, as he was careful enough to point out, does not mean that they have nothing to do with the state of affairs in our world: "from thought experiments most people learn about their concepts and the world together" (253).

[26] Popper, *Logic,* 444, demands that idealisations are "concessions to the opponent or at least acceptable to the opponent"; Rescher, *What if,* 9-14, insists from the outset on the need for a context.

will use simple examples to make clear that when we eliminate one of these conditions, we have a thought experiment that does not work. We will thus successively eliminate the condition that what a thought experiment tests are beliefs, the condition of agreement as to beliefs, and what is even more interesting in our minds, the condition of a hierarchy of beliefs. As above, we chose examples written at different periods and taken from various disciplines in order to point out that periods and disciplines are not pertinent parameters in our argument.

What if we eliminate the condition (I) that what we are testing are beliefs?

We will illustrate this case with the Cartesian thought experiment of receding bodies. Its domain is what Descartes calls "general physics", the part of physics that gives *a priori* demonstrations about matter and motion. In the article of the *Principia philosophiae* where the thought experience is carried out, Descartes wants to prove that the nature of matter, that is of bodies in general, is to be an extended substance. He must consequently eliminate all the qualities that we experience in bodies. The thought experiment concentrates on the difficult case of hardness and aims at refuting the atomist claim that hardness belongs to the nature of bodies.

> As for hardness, the only thing that our senses indicate to us, is that the parts of hard bodies resist the motion of our hands when they touch them; however, if whenever our hands moved in some direction, all bodies existing there receded with the same speed as our hands approached them, we would never feel hardness. But it is impossible to admit that bodies that recede in that way would for that reason lose the nature of bodies. By the same argument it can be shown that weight and colour and all other qualities of that sort that we sense in a material body can be taken away from it, leaving it intact.[27]

In order to understand how Descartes deals with this problem, and what makes his thought experiment somewhat queer, let us first make explicit his system of beliefs:

    (1) We are in the A-world.

    (2) In the A-world, we touch bodies.

    (3) When we touch bodies, resistance is felt.

---

[27] *Principia philosophiae*, II 4, Descartes, *Œuvres*, VIII/1: 42, our translation.

(4) A body that offers resistance to motion is a hard body.

(5) Hardness belongs to the nature of bodies.

(6) We have knowledge about the nature of things in general (and bodies in particular).

(7) The nature of X is what X cannot be conceived without.

(8) The nature of bodies is independent from what we can feel about bodies.

In this consistent set, there are quite different kinds of beliefs. (4) and (7) are definitions, the first one derived from a phenomenological experience of a physical property, the last one specifying the meaning of a metaphysical term. As many metaphysical definitions, (7) has an unquestionable modal content, that makes this thought experiment undeniably modal. However, as we explained earlier, we are not so much interested in identifying types of thought experiments, which would imply relying on modal logic, than in showing how a thought experiment works. (2)-(4) are derived from our experience of bodies. (6)-(8) are metaphysical statements. To carry out the Cartesian thought experiment is to introduce the negation of (1) and (2): we emigrate to a W-world, in which, contrary to what happens in our A-world, we do not touch bodies because they recede from us. What happens in that case? Would we say that hardness belongs to the nature of bodies? It could be expected that, in a W-world, we would have other beliefs about bodies, for example, we can imagine that W-hardness would be, for W-people, the rate of speed with which bodies recede from their hands. If it were the case, the experience of W-bodies in the W-world would lead W-people to a different definition of hardness, so that (5) would be perhaps conserved, but with quite another meaning, tied to the different experimental contributions that would be expressed in the form of beliefs (W-3) and (W-4). However, Descartes claims that, in the W-world, we keep our beliefs (3) and (4), notwithstanding the fact that they are derived from our experience of A-bodies. Consequently, his test of consistency and his procedure of prioritisation of beliefs operate only on the subset (5)-(8). Now, for him, (6)-(8) are metaphysical statements that take precedence over (5): it is less costly to abandon (5) than to abandon (6) — which means being a sceptic as to the knowledge of the nature of bodies — or (7) — which means giving up a well-established definition of what is a nature — or (8), that can be seen as a consequence of (7).

The receding bodies thought experiment would most probably not have been accepted by Descartes' opponents, the 17th century atomists. They would have contested (4), because what they attribute to bodies was not the kind of phenomenological or sensible hardness Descartes talks about here, but an essential impenetrability defined as the property of a body to prevent any other body to occupy its place. However, it is another question that we would like to raise here — namely the question of why Descartes believes that in the W-world, beliefs (2)-(4) would not significantly change. It is obviously because his goal is to examine the nature of a body. Inasmuch as the nature of X is defined by (6) that without which X cannot be conceived, examining the nature of X means declaring the condition of possibility not of X, nor even of our experience of X, but of our conception of X. In the case of bodies, this means setting aside everything that is felt (for feeling is not conceiving), and keeping that without which neither bodies nor their properties can be conceived, that is extension.

If the Cartesian thought experiment had only the goal of defining the conditions under which bodies can be conceived, in other words, to use the terminology generally adopted in this article, making explicit the conditions under which we can have beliefs concerning bodies, we would stop at this point and say that we cannot conceive of bodies or of their properties, potentially sensory properties, independently of extension. However, a few lines earlier, at the beginning of Book II, Descartes warned us. The fact from which one should start is the existence of an extended substance that has all the properties that are clearly perceived to belong to an extended substance, this extended thing being what is usually called body or matter.[28] Hence, what Descartes means to speak of in Book II is not only the nature of bodies, but the bodies that exist; consequently, when he presents the receding bodies thought experiment, while the question is one of eliminating hardness and other sensory qualities, it is with reference to bodies that exist in our world. In other words, Descartes plays both sides of the fence: his thought experiment allegedly teaches us something both about the conditions under which a body can be conceived, and about the bodies of our world.

Indeed, this is a general problem in Descartes, not a problem specific to the receding bodies thought experiment. But we think that it is particularly striking in this

---

[28] *Principia philosophiae*, II 1, Descartes, *Œuvres*, VIII/1: 41.

case and that it is a nice illustration of the fact that we go off track if we drop the condition that thought experiments are exercises in reasoning about our beliefs and replace it with the idea that they are about the things themselves.

What if we eliminate the condition (II) of agreement on beliefs?

In general, thought experiments help us clarify our belief systems, revealing gaps or avoiding inconsistencies in them. Being exercises in reasoning, they have their strong points, but their weak points as well. In particular, it can happen that thought experiments fail to hit on a shared belief, be it a common experience or more generally a shared assumption. This might have been the case if Salviati had conversed not with Simplicio, but with an Aristotelian who would have refused the assumption that speed is mediative. This might have been the case as well if Thomson's opponents had been unable to agree with her on the language of rights. In these imaginary cases, Salviati's and Thomson's opponents might have begun to contrive counter-thought experiments and might have failed to find common ground. Examples of this kind of an out-of-control thought experiments-machine may be found in most controversies, but are especially striking in what Norton aptly called the thought experiment/anti-thought experiments pairs. As already noted by others, this is the case of the two Marys, brilliant scientists who, from birth, were confined to a laboratory, where they had only black and white experiences, but learnt everything about colours and colour-perception. Mary Jackson experiences something new when she leaves her laboratory and sees red for the first time; Mary Dennet however, when she leaves her lab, just exclaims: "Ah! Colour perception is just as I thought it would be!"[29] Exchanges of such thought experiments will quickly look like a fool's game, because they do not constitute *per se* a reason to choose one clear, complete and consistent system of beliefs rather than the other.

What if the condition (III) of hierarchy of beliefs is eliminated?

We will illustrate this case by an argument of counterfactual style that comes from Aristotle's *Politics*:

---

[29] Norton, "Why Thought Experiments," 45-49; Atkinson and Peijnenburg, "Thought Experiments Poor," 308-315.

> […] if every tool could perform its own work when ordered, or by seeing what to do in advance, like the statues of Daedalus in the story, or the tripods of Hephaestus which the poet says 'enter self-moved the company divine' – if thus shuttles wove and quills played harps of themselves, master-craftsmen would have no need of assistants and masters no need of slaves.[30]

We will pass over quickly two preliminary difficulties. The passage quoted follows a formula that presents a translation issue; Aristotle has just spoken of the slave as an *organon pro organôn*. The formula can mean "a tool which serves for several tools" or else "an instrument that comes before the other instruments". According to the first interpretation, Aristotle wanted to oppose the specialisation of the instrument to the polymathism of the person using it; according to the second, he wanted to oppose the passivity of the instrument to the activity of the person using it. A second difficulty is that we cannot put on the same plane shuttles and harps. The first are instruments of *poiêsis*, the second instruments for the *praxis*: they thus refer to extremely different types of action. The interpretation that follows shows how we have chosen to resolve this conflict. Let us ask ourselves: "what are Aristotle's beliefs that the counterfactual supposition is supposed to undo?" Let us admit that the harp here is emblematic of instruments belonging to the order of the *praxis*. We would then have something like:

(1) An instrument is a thing able to accomplish the work that is proper to it.

(2) For an instrument, accomplishing the work that is proper to it, is to operate.

(3) Because it is unable to move itself spontaneously and even less able to foresee what is expected of it, an instrument is unable to operate by itself.

(4) In order for an instrument to operate, there must be an instrument of instruments.

(5) When instruments of instruments are human beings,[31] they are, depending on the case, slaves or workers.[32]

(6) Slaves and workers are in the service of masters and craftsmen respectively.

---

[30] Aristotle, *Politics*, 1253b17. We thank Frans De Haas for his sharp and constructive criticisms of our first presentation of this example.

[31] These could also be animals, but we will set aside this point.

[32] Depending on whether one is dealing with instruments of *poiêsis* (such as shuttles) or of *praxis* (such as zithers). On zithers as instruments of *praxis*, see *Magna moralia*, 1997a9-15.

These are beliefs that Aristotle himself would use in a complex argument aimed at establishing that the master-slave relationship is natural. But let us consider them in and of themselves: they advance definitions [(1) and (2)]; they enlighten the relationship between craftsman and assistant or that between master and slave [(5) and (6)]; they tell us what are the negative [(3)] or positive [(4)] conditions that must be in effect for an instrument to be operational. Now it is clear that the thought experiment in this passage invites the reader to consider what would happen if the negative condition for the operability of instruments was modified. What would happen if the instruments became capable of moving on their own, or understanding what was expected of them? In other words, what would happen if there was a non-(3)? With the assumption that non-(3) is the case, (4) and (5) and (6) fall one after the other, like the legendary dominos: if an instrument can operate by itself, it is no longer indispensable that instruments of instruments make it operational, and it thus becomes futile to wonder what is the status of these instruments of instruments, whether in the field of *poiêsis* or in that of *praxis*. In short, the introduction of non-(3) does not allow us to identify a weak link that we could then eliminate: it jeopardises Aristotle's entire set of beliefs,[33] which is why he prefers not to reorder them and ends up refusing to admit non-(3) as absurd. The reason is that the hierarchy among these beliefs is not clearly determined: if one goes down, the other goes down with it: united they stand, divided they fall.

Conclusion

As a conclusion, we will make four remarks: first, on the incomplete character of our argumentation; second, on the domain of validity of our description of the mechanisms of thought experiments, with respect to different historical periods and different fields of knowledge; third, on our position in the contemporary debate on thought experiments; and fourth, on the fact that our position does not imply any kind of general relativism.

i) In the course of our argumentation, we argued through counter-examples that the three conditions (I), (II) and (III) were individually necessary, not that they were conjointly sufficient. To proceed by induction to demonstrate this would be tedious, not

---

[33] But not, however, (1) and (2), which give prior definitions of the instrument and the operation that remain valid, whether the instruments can move themselves or not.

to say impossible, and our paper does not meet the conditions for a rigorous proof — in particular because, even if we admitted the broad characterisation of thought experiments advanced in the introduction of this book, we do not have a clear-cut definition of the notion. It is thus only as a conjecture that we will advance the thesis that the conditions (I)-(III) are conjointly sufficient, a thesis we will hold to be true until a counter-example can be opposed to it.

ii) The conditions (I)-(III) are conditions on beliefs that are not proper to a historical period or a field of knowledge.[34] But we can understand retrospectively why they are easily met in the early mechanics of the 17th century analysed by Mach: operations on bodies, some more elementary than some others, had led to the constitution of a reservoir of shared and hierarchised beliefs. We also understand that thought experiments are usually effective in a field like physics, where there are principles or laws upon which every physicist agrees, and which trump all other propositions. We finally understand what happens in these out-of-control thought experiments-machines that sometimes emerge in physical controversies. But we do not see why there would be a sharp essential difference between thought experiments in the hard sciences and in the other fields of knowledge. There is no *a priori* reason for our conditions not to be met elsewhere: this is merely a *de facto* question. They were met in the case of Thomson's thought experiment dealing with beliefs related to propositions concerning rights: this does not mean that it is the case in the whole domain of applied ethics. Thought experiments appear much more difficult in the present state of economics than in physics, but this state of affairs does not rule out the existence of thought experiments in a particular economical school or, in the future, in economics in general.[35]

iii) With respect to the existing literature, the description we just proposed of thought experiments inclines us, without committing us, to a deflationary position in at least two respects. As we noted to begin with, our description of the mechanism of

---

[34] This was already a background assumption of Atkinson and Peijnenburg, "Thought Experiments Poor". On the contrary, Wilkes, *Real People,* and Rescher, *What if,* argue that there is an essential difference between philosophical and scientific thought experiments.

[35] Cartwright, "Vanity of Rigor," assimilates theoretical models in economics and thought experiments in physics, because both "seem like toys. They have no texture to them, no tones of colour, no gradations. They are simple situations, with caricature objects with few options for their behaviour" (217-218). Her argument, quite different from ours, relies on a confusion between idealisation and thought experimentation.

thought experiments does not commit us to any position concerning their nature. In particular, we do not need the strong deflationist thesis that thought experiments are only arguments; it is sufficient for us to admit that they are arguments. However, we have to confess that we have more sympathy for the deflationist thesis of Norton, according to which thought experiments are only arguments, albeit with some qualifications, than for the imaginarist thesis of Gendler that a thought experiment contains something more than an argument. Even so, our aim was not to defend the deflationist thesis as such, and it is only as a conjecture that we will advance the thesis that thought experiments are arguments, a thesis we will hold to be true until counter-examples can be opposed to it.

There are mainly two reasons for which we are not convinced by Gendler´s thesis. First, we think that she was induced in the first place to defend such a thesis because of her excessive expectations with regard to arguments and, more generally, because of her most unrealistic vision of the sciences. In her paper on Galileo, she writes, for example, that "to draw a conclusion on the basis of a *non-thought-experimental argument* is to be led by a process of inductive or deductive reasoning from a set of explicit premises".[36] Her argumentative strategy is therefore the following: inasmuch as the strapped-bodies thought experiment is not totally convincing without "broad, defeasible, tacit assumptions", that are not explicit in Galileo's dialogue, something should stand in for them.[37] But, contrary to what she implies, no real experience in physics would be convincing, if background knowledge were not mobilised. This does not make physical experimentation mysterious: a certain number of these implicit premises are made explicit when they are contested, for example during controversies. Second, Gendler does not succeed in clearly specifying what would be this supplement that would help understand the outcome and justification of a thought experiment. She speaks of "quasi-sensory intuitions" or "mental images" and invokes acts of cognition that do not result from explicit calculating or reasoning, like imagining how many elephants can fit in a room, or whether this carpet would clash with bright green painted walls.[38] Such acts of cognition indeed exist, as the literature she referred to amply shows,[39] but it is not at all

---

[36] Gendler, "Galileo," 399.
[37] Ibid., 406-408. For a detailed criticism of Gendler's position, see Zeimbekis' essay in this volume.
[38] Gendler, "Thought Experiments Rethought," 1156-1159.
[39] Ibid., 1159-1161.

certain that they are actually involved in thought experiments, which do not appear on the mental scene of an individual, but in highly sophisticated controversial contexts, where a choice between competing hypotheses or even competing theories is to be made.[40]

We consequently, if only because of a general principle of economy of thought, feel closer to the deflationary thesis defended by Norton. Note however that Norton, inasmuch as he was answering Brown,[41] explicitly links together his deflationary thesis and his empiricism: if thought experiments were not merely arguments, empiricism would be endangered by "a mysterious route to knowledge of the physical world", "epistemic miracles" or "epistemic magic".[42] While we are ourselves more empiricists, — although we would have to explain in which sense —, we think it is important to note that the reciprocal proposition (empiricism implies the thesis that thought experiments are merely arguments) does not seem to us to be assured. The analysis inspired by Rescher that we have proposed to specify the mechanism of thought experiments seems to us to be compatible with all sorts of positions in terms of theory of knowledge, not only with empiricist positions. At the risk of repeating ourselves, materials providing input in thought experiments come out of beliefs that can have all sorts of origins.

iv) Speaking of "shared beliefs" can evoke certain relativistic theses, if only because this expression is sometimes used by the tenants of the thesis that a shared belief is *per se* a good belief. Our position with respect to relativism is simple: we are contextualists as to thought experiences, but we do not feel that this local contextualism has any reason to lead to relativism, not to speak of general relativism.

We can perhaps make ourselves understood by a little thought experiment of our own — which, at this point of our paper, is rather an illustration. Let us suppose that a group of individuals share the following beliefs: there is a very bad being; this very bad being has three and only three hairs; of these three hairs, one is blue, the other yellow, the last one red. Suppose moreover that these three beliefs are assumptions so fundamental for this group of individuals that they have developed, without ever doubting them, fairly sophisticated theories based on them. If these individuals enjoy arguing as we do, they

---

[40] Norton, "Why Thought Experiments," 60-61, criticises on similar grounds the mental modelling views of thought experiments.
[41] Brown, "Thought Experiments"; Brown, "Why Thought Experiments Transcend," 30-31.

have a set of beliefs that meet the three conditions we have defined: they will thus be able to put in place thought experiments that work, thanks to which one of them can convince others of some particular property of this very bad being. But now one of the individuals of this group, not only talkative but adventurous, leaves her companions to meet with a very different group: will she not perceive that the thought experiments she has laboriously put in place no longer work in this new group? This illustrates that a thought experiment does not work in and of itself, but in a certain context, but this does not call on a relativist thesis as such, but only on a minimal contextualism. Relativism does not consist in saying that there are shared beliefs in certain contexts, beliefs that are not shared in others; it is the thesis that shared beliefs, because they are shared, are good beliefs, and in particular that there is no epistemic criterion for choosing among opposing and competing beliefs. In other words, for a relativist, the best belief is the most widely shared belief, the nature of this 'best' being consequently variable. In our paper, we take no position as to this relativist thesis.

However, even if we were relativists with respect to thought experiments, we do not see why this local relativism would necessarily imply a general relativism, extended to the physical sciences. What keeps us from being relativists in physics is not the conviction that there would be, if only once, a pure observation, a true law of nature, a concept corresponding to a natural kind: we know how problematic each of these affirmations is. It is, to say it naively, the fact that there are physical theories which, taken as a whole, work, insofar as, and without us needing to determine here why, they allow for a certain number of predictions, which is today generally not the case, for example, in history, in economics, in philosophy or in sociology. This is the reason why certain thought experiments have a predictive character in physics; their predictability coming not from the thought experiments as such, but from the fact that they occur within the context of predictive theories.[43] Thus, even if we were not only contextualists, but were

---

[42] Norton, "Thought Experiments," 129; Norton, "Are Thought Experiments," 334-337; Norton, "Why Thought Experiments," 44-45.

[43] For a similar move, see Atkinson and Peijnenburg, "Thought Experiments Poor". They note, however, that unsuccessful thought experiments are less disastrous in science than in philosophy because, in science, we can turn to theories (and there are good reasons to prefer one theory over the other) or to real experiments (and there are crucial experiments in the sciences).

relativists with respect to thought experiments, our relativism would *prima facie* not be extended to all the physical sciences.

Bibliography

Aristotle: *Politics*, translation by H. Rackham (Cambridge, MA, 1932).

———. *Metaphysics X-XIV, Oeconomica, and Magna Moralia*, translation by H. Tredennick and G. Cyril Armstrong (Cambridge, MA, 1935).

Atkinson, David: "Experiments and Thought Experiments in Natural Sciences", in *Observation and Experiment in the Natural and Social Sciences*, 232, ed. Maria Carla Galavotti (Dordrecht, 2003), 209-225.

Atkinson, David and Jeanne Peijnenburg: "When Are Thought Experiments Poor Ones?", *Journal for General Philosophy of Science,* 34 (2003), 305-322.

———. "Galileo and Prior Philosophy", *Studies In History and Philosophy of Science* (2004), A/35: 115-136.

Brown, James Robert: *Laboratory of the Mind: Thought Experiments in the Natural Sciences* (London: 1991).

———. "Thought Experiments: a Platonic Account", in *Thought Experiments in Science and Philosophy*, ed. Tamara Horowitz and Gerald J. Massey (Savage, MD, 1991), 119-128.

———. "Why Thought Experiments Transcend Empiricism", in *Contemporary Debates in the Philosophy of Science*, ed. Christopher Hitchcock (Oxford, 2004), 23-43.

Cartwright, Nancy: "The Vanity of Rigor in Economics: Theoretical Models and Galilean Experiments", in *Hunting Causes and Using Them. Approaches in Philosophy and Economics* (Cambridge, 2007), 217-236.

Descartes, René: *Œuvres de Descartes*, ed. Charles Adam and Paul Tannery, new pres. by B. Rochot and P. Costabel (11 vols.), (Paris, 1964-1974)

Duhem, Pierre: *La théorie physique* (Paris, [1906] 1981).

Galilei, Galileo: *On Motion and On Mechanics: Comprising De Motu* (ca. 1590), translated with Introduction and Notes by I. E. Drabkin and *Le Meccaniche* (ca. 1600), translated with Introduction and Notes by Stillman Drake (Madison, 1960).

———. *Two New Sciences*, ed. and tr. by Stillman Drake (Toronto, 1989).

Gendler, Tamar Szabó: "Galileo and the Indispensability of Scientific Thought Experiment", *The British Journal for the Philosophy of Science,* 49 (1998), 397-424.

———. "Thought Experiments Rethought — and Reperceived", *The British Journal for the Philosophy of Science*, 71 (2004), 1152–1163.

Häggqvist, Soren: *Thought Experiments in Philosophy* (Stockholm, 1996).

Hitchcock, Christopher (ed.): *Contemporary Debates in the Philosophy of Science* (Oxford, 2004).

Horowitz, Tamara and Gerald J. Massey (ed.): *Thought Experiments in Science and Philosophy* (Savage, MD, 1991).

Jamieson, Dale: "Method and Moral Theory", in *A Companion to Ethics*, ed. Peter Singer (Oxford, 1991), 476-490.

Janis, Allen: "Can Thought Experiments Fail", in *Thought Experiments in Science and Philosophy*, ed. Tamara Horowitz and Gerald J. Massey (Savage, MD, 1991), 113-118.

Koyré, Alexandre: "Le *De Motu gravium* de Galilée. De l'expérience imaginaire et de son abus", in *Études d'histoire de la pensée scientifique* (Paris, 1973), 224-271.

———. "Pascal savant", in *Études d'histoire de la pensée scientifique* (Paris, 1973), 362-389.

Kuhn, Thomas: "A Function for Thought Experiment", in *The Essential Tension* (Chicago, 1977), 240-265.

Kühne, Ulrich: *Die Methode des Gedankenexperiments* (Frankfurt, 2005).

Mach, Ernst: *The Science of Mechanics*, trans. Thomas J. McCormack (La Salle, IL, 1960)

———. *Knowledge and Error. Sketches on the Psychology of Enquiry,* trans. Thomas J. McCormack and Paul Foulkes (Dordrecht, 1976).

Norton, John D.: "Thought Experiments in Einstein's Work", in *Thought Experiments in Science and Philosophy*, ed. Tamara Horowitz and Gerald J. Massey (Savage, MD, 1991), 129-148.

———. "Are Thought Experiments Just What You Thought?", *Canadian Journal of Philosophy*, 26 (1996), 333-366.

———. "Why Thought Experiments Do Not Transcend Empiricism", in *Contemporary Debates in the Philosophy of Science*, ed. Christopher Hitchcock (Oxford, 2004), 44-66.

Pabst Battin, Margaret: *Ethical Issues in Suicide* (Prentice-Hall, NJ, 1995).

Popper, Karl: *The Logic of Scientific Discovery* (London, 2002).

Rescher, Nicholas: *What if? Thought Experimentation in Philosophy* (New Brunswick, 2005).

Sorensen, Roy: *Thought Experiments* (Oxford, 1992).

———. "Thought Experiments and the Epistemology of Laws", *Canadian Journal of Philosophy*, 22-1 (1992), 15-44.

Thomson, Judith Jarvis: "A Defence of Abortion", *Journal of Philosophy and Public Affairs*, 1 (1971), 47-66.

Wilkes, Kathleen: *Real People: Personal Identity without Thought Experiments* (Oxford, 1988).

Index