Thought Experiments as a *sui generis* scientific tool

A scientist's toolbox is multiple and diverse, it contains: experiments, computer simulations, arguments, etc. One of these tools, thought experiments (hereafter TEs), have become a hot topic in philosophy, especially due to Kuhn's (1964) puzzling question: "How […] relying exclusively upon familiar data, can a thought experiment lead to new knowledge or to new understanding of nature?" Part of the current debate on TEs, as it has been formulated by the two protagonists (*i.e.* Norton and Brown, 2004) addresses the question: "Do TEs transcend empiricism?"

For Norton (since 1991) they *do not.* He defends an *argument account* of TEs and argues that the only ''non-miraculous'' way to get new information about the world without resorting to new empirical data is through arguments from premises describing past empirical data about the world. Thus, TEs are just "disguised" deductive or inductive arguments with *irrelevant* and *eliminable* "particulars". TEs are thus dispensable and a "successful" TE is just a "good" logical argument.

While for Brown (since 1986) TEs *do* transcend empiricism. He regards them as powerful counterexamples to empiricism and defends a *Platonic intuition-based account* of TEs. He argues that some TEs, those that he calls "platonic" (e.g. Galileo's Pisa tower), involve no new empirical data and are not reducible to logical arguments. He urges that platonic TEs take us beyond old empirical data and provides us with new *a priori* knowledge about the world (*i.e.* Galileo's law of free fall). Though TEs are instances of *a priori* reasoning, they are still fallible for Brown; briefly a platonic TE fails if our intuition, or "platonic perception", fails. Beyond Brown's analogy between the fallibility of our ordinary perception of concrete objects, such as rabbits and stones, and the fallibility of our platonic perception of abstract entities, such as mathematical objects and laws, it remains mysterious how this latter succeeds or fails.

This debate has generated an extensive literature which provides a middle ground between Norton's empiricism and Brown's rationalism. TEs have been defined within different accounts, such as the following: "*mental models"* (e.g. Nersessian 1993; Cooper 2005) "*experimentalist"* (e.g. Sorensen 1992, Lennox 1991, Buzzoni 2010), "*constructivist"* (e.g. Gendler 1998) and *"intuition-based"* (e.g. Brendel 2004). Despite their differences, the common ground among most of these accounts concerns the notion of possibility at play in the scenario of a successful TE: the requirement is that the scenario of a successful TE should be *realisable in principle – i.e.* nomologically possible; that is possibility under actual theory, law and principle. If such a requirement does not hold, the TE is bound to fail (*cf.* El Skaf 2017).

In spite of this proliferation of epistemic accounts, many unresolved questions persist, for instance: Is it possible, and even desirable, to articulate a precise and complete definition of what TEs are? What is the nature of the new knowledge they bring? What is the nature and role of the elements – "data" (Kuhn 1964) "particulars" (Norton 1991) or "state of affairs" (Gendler 2000) – imagined in the scenarios of TEs? Are there relevant and irrelevant elements and how can we recognize which are which? What notion of possibility, as opposed to the actuality of REs, should we expect to deal with in TEs? Finally, which are the TEs' cognitive underpinnings? Are they propositional, non-propositional, or a mix of both?

The literature is thus characterised with a multitude of accounts, which results in wide divergences as to what TEs are, how they function and how they justify their conclusions. In addition, most accounts rely on *a-historically* analysed case studies, which yields disagreements about the conclusions we can derive from some TEs, and thus divergences pertaining to their epistemic function. Worst, this lack of historical analysis sometimes turns a philosophical debate into a red herring: the ongoing epistemological debate on TEs revolves, in part, around Galileo's Pisa tower TE, in particular how it justifies its conclusion. Nevertheless, the TE's function is misrepresented in part of the philosophical literature as *revealing* and *justifying* Galileo's law of free fall (e.g. Brown since 1986). While for Galileo instead, the TE's function, *per se*, was to *refute* Aristotle's law and was part of two "argumentative strategies", excogitated by Galileo precisely to defend two different laws of free fall, in 1590 and then in the 1630'.

Finally, most empirically oriented accounts found in the literature are *either* too restrictive – TEs are required to be realisable in principle (*i.e.* nomologically possible) – *or* reductive – TEs are reduced to other scientific tools such as arguments, computer simulations, mental models or real experiments – *or even both* restrictive and reductive.

In this paper, I defend a novel non-reductive, non-restrictive epistemic account of scientific TEs, compatible with empiricism and built on case studies from the history of physics. First, I survey the recent history of the debate on TEs. Second, I formulate this account's two main claims; *i.e.* (i) TEs are *inconsistency revealers and eliminators* and (ii) TEs share a *common general structure*. Third, I provide illustrations with case studies taken from the history of physics. Fourth, I expose 4 ways TEs could fail, which will allow me to further formulate this account along the way. Finally, I conclude with a brief comparison with several accounts of TEs found in the literature.