Novelty Journals

ISSN 2394-9694

International Journal of Novel Research in Humanity and Social Sciences
Vol. 2, Issue 2, pp: (1-3), Month: March - April 2015, Available at: www.noveltyjournals.com

Counterfactual Truths: The Logical Structure of Argumentative Thought Experiments

Javier Y. Álvarez-Vázquez

Department of Philosophy, Heidelberg University, Heidelberg, Germany

Abstract: **Argumentative thought experiments are structurally conditional clauses. They can hence be formalized by means of the principle of modus ponendo ponens, as well as of modus tollendo tollens. In contrast to the practice in formal logic, exponents of argumentative thought experiments claim that the logical validity of a conclusion drawn within the framework of a particular conditional argument also holds beyond the particular conditional in question. In this paper, I articulate the criticism that this claim is wrong by arguing that the counterfactual scenario sets itself the most determinant premise. If the counterfactual scenario sets the initial conditional premise of the argument, then its true conclusion holds only as a counterfactual truth. The present paper illustrates this criticism using Frank Jackson's thought experiment, the so-called knowledge argument, as a concrete example.**

Keywords: **thought experiments; counterfactual scenario; truth; knowledge argument; Frank Jackson; modus ponendo ponens; modus tollendo tollens; validity; valuation.**

I. INTRODUCTION

The *argumentative* type of philosophical thought experiments is often used fallaciously.[1] The fallacious practice consists in the treatment of deduced valid conclusions as universal or unconditioned true conclusions. This erroneous practice is closely related to the function which argumentative thought experiments are intended to fulfill. In contrast to the argumentative type, *illustrative* philosophical thought experiments are, for instance, used to clarify and exemplify an argument or an idea that does not depend, either in content or in structure, upon the thought experiment itself. *Experimental* thought experiments, like Einstein's famous train experiment [2], also contrast with the argumentative type. Experimental thought experiments are usually intended, for whatever reason, to substitute a real experiment, which might be conducted later. Unlike their experimental and illustrative counterparts, argumentative thought experiments are intended to serve as an argument themselves. The narrative of argumentative thought experiments simultaneously provides, in addition to a depiction of a counterfactual scenario, the elements of a logical argument. Because of their counterfactual character, argumentative thought experiments are *per definitionem* conditionals; they are—so to speak—*if-arguments*. It is in this aspect where the erroneous practice comes about. While the validity of a conditional relies on its syntax or structure, the value of its truth additionally depends on its semantic. The fallacious practice of philosophers that use this kind of thought experiments resides, therefore, in the identification of the truth-value with the validity of the argument.

By reference to Frank Jackson's prominent thought experiment of the brilliant neuroscientist Mary [3], I demonstrate in Section 2 how this fallacious practice exemplarily takes place. In Section 3, I draw the conclusion directly from the concrete example that the truths of argumentative thought experiments are not universal but rather counterfactual truths. This conclusion does not rely, as I shall demonstrate, on the syntax alone, but furthermore on the semantic which the argumentative thought experiments entail.

[1] For a typology of thought experiments, see [1], p. 74ff.

International Journal of Novel Research in Humanity and Social Sciences
Vol. 2, Issue 2, pp: (1-3), Month: March - April 2015, Available at: www.noveltyjournals.com

II. DEMONSTRATION

In order to concentrate the discussion on the particular problem I want to demonstrate, I will leave aside all other problems related to Frank Jackson's thought experiment.[2] Problems with the picture that Jackson is trying to set up, for instance, will be here completely ignored. In this line of thinking, we want to take for granted that the counterfactual scenario of a color-blindness-like environment (acromatopsia) is narratively well achieved. We take also as given that there is agreement upon the understanding of concepts like 'knowledge', 'learning' or 'having all physical information'. Finally, we want to assume that Jackson's thought experiment succeeds in its intention.

Given all these assumptions, I want to demonstrate that the fallacious practice of attributing a universal truth-value to counterfactual truths does not rely upon the logical argument in question. Rather, the fallacy extends beyond the counterfactual character itself. Even assuming that Jackson's argumentative thought experiment succeeds in its intention, it proves itself to be fallacious by claiming its truth as universal or unconditioned. Our example reads in its essential passages as follows:

> Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room *via* a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, of the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces *via* the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue'. […]
> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. *Ergo there is more to have than that, and Physicalism is false*. ([3], p. 130. My emphasis.)

This argumentative thought experiment can be formalized after the principle *modus ponendo ponens.* It is intended to prove that physicalism is false, where physicalism is exemplified as 'Mary does not learn anything new'. The implied sentences are defined as follows:

$V_1(p)$ = Mary live in a color-blindness-like environment and has all physical information (knowledge) concerning colors and color-related phenomena.
$V_1(q)$ = Mary leaves the color-blindness-like environment and experiences the world in its full-color appearance.
$V_1(r)$ = Mary does learn something about colors or color-related phenomena from that experience.

Given this valuation of the sentences, we can formalize it as the following logical argument:

$$(p \wedge q) \rightarrow r$$
$$p \wedge q$$
$$\overline{\qquad\qquad}$$
$$r$$

From the viewpoint of a "qualia freak", as Jackson describes himself, and as the structure of the conditional refers, the occurrences of p and q are sufficient conditions for r. Although the possibility of other antecedents for the occurrence of r remains open, those possible antecedents have to be mentioned in order to build a logical argument with them. Such other possibilities are absent in Jackson's narrative and therefore they have to remain unknown for the occurrence of r.

[2] For an extensive discussion on the many problems related to Jackson's thought experiment, see [4].

However, for a physicalist, who might insist that Mary does not really learn anything new in spite of the new lived experience, the occurrences of p and q are more than just sufficient conditions. Jackson's thought experiment exemplifies the attacked physicalism as the case that 'Mary does not learn anything', namely $\neg r$. Given the case that Mary does not really learn anything new ($\neg r$) and the conditional still holds, the argument reads as follows:

$$\begin{array}{l} (p \wedge q) \rightarrow r \\ \neg r \\ \hline \neg\,(p \wedge q) \end{array}$$

This formalization of the argument implied in Jackson's thought experiment is in fact the *modus tollendo tollens* version of the conditional written above. It characterizes the initial premise that expresses the occurrences of p and q as a necessary condition for r. In contrast to the *modus ponendo ponens* version, where the initial premise is characterized as a sufficient condition, here it receives a stronger characterization as a *necessary* condition in relation to the case that $\neg r$.

III. CONCLUSION

Based on these logical facts and argumentative dynamics, I argue that if r is true, its truth-value either depends sufficiently or necessarily from the initial premise of the conditional. In the case of argumentative thought experiments, as discussed above, their initial premises are in their content constituted by the counterfactual scenarios. This means that argumentative thought experiments, insofar understood as conditional logical arguments, are always conditioned by their counterfactual scenarios. In contrast to the predominant treatment of the consequences from such argumentative thought experiments, the truth deduced from their premises is also conditioned by their counterfactual framework. The truths of such counterfactual arguments are, therefore, counterfactual truths. As counterfactual truths, they do not hold as universal or unconditioned, because their truth-value depends on the occurrences of p and q, as well as their particular valuation.

REFERENCES

[1] D. Cohnitz, Gedankenexperimente in der Philosophie. Paderborn: Mentis, 2006.

[2] A. Einstein, Über die spezielle und die allgemeine Relativitätstheorie. Berlin: Springer, (1916) 2009.

[3] F. Jackson, "Epiphenomenal Qualia," The Philosophical Quarterly, vol. 32, pp. 127-136, April 1982.

[4] P. Ludlow, Y. Nagasawa, and D. Stoljar (Eds.), There's Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument. Cambridge, Mass.: MIT Press, 2004.