The *Challenge of the illusion*: Understanding defective theories

María del Rosario Martínez-Ordaz, Moisés Macías-Bustos

Understanding has been traditionally considered to "consist of knowledge about relations of dependence. When one understands something, one can make all kinds of correct inferences about it" ([1]: 100). In addition, understanding is often regarded as factive, this is, the content of understanding can only include true propositions. This considered, having the impression of understanding something that is knowingly imprecise, incoherent or false is called the 'illusion of depth of understanding' (see [1], [2]). While such an illusion is clearly a mental confusion, according to some, certain falsehoods can facilitate understanding if and only if they provide scientists "with epistemic access to true information that is difficult or even impossible to discern otherwise" ([3]: 26) −only in virtue of the true elements they shed light on.

Nonetheless, much scientific practice makes use of defective (conflicting, inconsistent, false and even impossible) theories; some of the most famous examples of this are: the early calculus, Bohr's theory of the atom and Frege's foundations of arithmetic, among others (see [4], [5],[6], [7], [8], [9]). And despite the fact that some of these theories are knowingly defective, scientists still report having 'understood' both the theories as well as the phenomena that they describe; these reports are often considered to be illusions. The demand for explaining how these reports can be genuine is what we call here the *Challenge of the illusion*.

There are broadly two very general ways to go about explaining what is going on when we grasp or communicate some defective theory: either claim that in none of those cases do we understand the aforementioned theories, which is on its face exceedingly implausible, or to say that we do understand them and then offer an explanation about what understanding comes down to in those cases. For this a strategy similar to the one already employed by structural realists for salvaging the continuity and preservation of science in light of pessimistic meta-induction style arguments may be adapted.

Part of the appeal of scientific realism is the claim that this view can explain why more mature scientific theories are more successful than their predecessors. However, on the one hand there are substantial changes in the ontologies and explanatory relations between any pairs of predecessor-successor theories, even those very close in time. But among the older theories with very different ontologies and different relations between the objects in the theory. Among theories we now consider false there are those which are strikingly successful, e.g. Newtonian mechanics, so it's not clear why, assuming our newest theories are also successful, their success is explained via their truth, since the earlier cases it was not. The structural realist can explain why reference is irrelevant for success, just because, in the intuitive way we just discussed, the theory has not assigned the right structure to the world. Furthermore it explains why a successful theory does not need genuine reference: it does not matter if the relation relating the terms is some specific relation R or if the property had by the terms related by R is some specific property P, the abstract description in terms of some set of objects in the domain (the objects having P) and some relation relating these objects in the right way will suffice to deliver truths in as much as the domain instantiates this structure (see [10], [11], [12]).

Analogously, we want to say something similar when it comes to understanding. What is understood in cases of defective theories is, broadly speaking, that some structure is being posited of some objects in some domain for the purposes of saying explanatory things about them given the posited structure. This works in cases where the theories are contingently false, since we can consider some possible structure instantiating the pattern with the ontology of the theory just so related and more importantly, where the theories are necessarily false: for example, whenever they are inconsistent and their underlying logic is classical i.e. they are explosive. For those cases we can consider impossible structures: where the ontology of the theory would be related in some patterned way if it were not for the inconsistent elements. For the purposes of this discussion, we focus on inconsistent theories and the case of Frege's foundations of arithmetic.

Here we content that when scientists report having understood a defective theory, even if clearly false or impossible, their claim might be legitimate. We argue that scientists understand a defective theory if they can recognize the theory's underlying pattern(s) and if they can reconstruct and explain what is going on in specific cases of defective theories as well as consider what the theory would do if not-defective –even before finding ways of fixing it.

In order to do so, we proceed in four steps.

First, we characterize very broadly scientific understanding. We also describe cases of legitimate understanding as those that at least exhibit (i) knowledge about relations of dependence, (ii) exemplification ([13], [14], [15]), and (iii) can be indicative of progress in the corresponding discipline. Second, we characterize defective theories and we explain in which way they are distinct from felicitous falsehoods. Here we also introduce the *Challenge of the illusion*. Third, we present the structuralist strategy and the analogous approach for the case of understanding. Fourth, we illustrate our view with a case study: Frege's foundations of arithmetic. This is a known inconsistent theory given Russell's paradox. However we are rather capable of discussing how the construction of arithmetic plays out in the theory ([16]); the proofs in the theory and their adequacy; the role of Basic Law V; of Hume's Principle and so on. In addition, since the theory is classical it is also trivial, yet we can consider it and make assertions about it which are informative ([4]).

Word count: 895.

References

[1] Ylikoski, P. (2013): "The Illusion of Depth of Understanding in Science", in H. De Regt, S. Leonelli & K. Eigner (eds.) *Scientific Understanding: Philosophical Perspectives*, Pittsburgh University Press: 100-119.

[2] Ylikoski, P. (2017): *The Illusion of Depth of Understanding in Science*.10.31235/osf.io/qz7sg.

[3] Lawler, I. (2019): Scientific understanding and felicitous legitimate falsehoods. *Synthese*. 10.1007/s11229-019-02495-0.

[4] Batens, D. (forthcoming): "Adaptive Fregean set theory", *Studia Logica*.

[5] Bueno, O. (1997): "Empirical Adequacy: A Partial Structures Approach", *Studies in History and Philosophy of Science* 28: 585–610.

[6] Bueno, O. (2006): "Why Inconsistency Is Not Hell: Making Room for Inconsistency in Science", in *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi*, Cambridge Studies in Probability, Induction and Decision Theory, Cambridge University Press: 70–86.

[7] da Costa, N. & O.Bueno (1997): "Inconsistent Mathematics", *The Journal of Symbolic Logic* 62: 683-685.

[8] da Costa, N. & S. French (2002): "Inconsistency in science: A partial perspective", in Meheus (ed), *Inconsistency in Science*, Kluwer Academic Publishers, Netherlands: 105-118

[9] da Costa, N. & S. French (2003): Science and Partial Truth, Nueva York: Oxford University Press.

[10] Russell, B. (1927): *The Analysis of Matter*, London: Kegan Paul, Trench, Trubner; New York: Harcourt, Brace.

[11] Votsis, I. (2004): *The Epistemological Status of Scientific Theories: An Investigation of the Structural Realist Account*, Doctoral Dissertation, London School of Economics (University of London).

[12] Votsis, I. (2018): "Structural Realism and its Variants", in J. Saatsi (ed.), *The Routledge Handbook of Scientific Realism*, London: Routledge: 108-119.

[13] Elgin, C. Z. (2009): "Exemplification, Idealization, and Understanding" in *Fictions in Science: Essays on Idealization and Modeling, Mauricio Suárez* (ed.), London: Routledge: 77-90.

[14] Elgin, C. Z. (2011): “Making Manifest: Exemplification in the sciences and the arts” *Principia* 15, 399— 413.

[15] Elgin, C. Z. (2017): “Exemplification in Understanding”, in Grimm, S.R., C.Baumberger & S. Ammon (eds.) *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, Routledge: 76—91.

[16] Zalta, E. N. (2019): ”Frege’s Theorem and Foundations for Arithmetic”, in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* , URL =`https://plato.stanford.edu/archives/sum2019/entries/frege-theorem/`.