volume VII♦ **issue** 3
2016

Logos & Episteme

an international journal
of epistemology

Romanian Academy
Iasi Branch
♦♦♦
"Gheorghe Zane" Institute
for Economic and Social
Research

Contact address:

Institutul de Cercetări
Economice şi Sociale „Gh.Zane"

Iaşi, str.T.Codrescu, nr.2, cod 700481
Tel/Fax: 004 0332 408922
Email: logosandepisteme@yahoo.com

www.logos-and-episteme.proiectsbc.ro

ISSN 2069-053

Logos & Episteme

TABLE OF CONTENTS

RESEARCH ARTICLES

WHAT MAKES LOGICAL TRUTHS TRUE?

Constantin C. BRÎNCUŞ

ABSTRACT: The concern of deductive logic is generally viewed as the systematic recognition of logical principles, i.e., of logical truths. This paper presents and analyzes different instantiations of the three main interpretations of logical principles, viz. as ontological principles, as empirical hypotheses, and as true propositions in virtue of meanings. I argue in this paper that logical principles are true propositions in virtue of the meanings of the logical terms within a certain linguistic framework. Since these principles also regulate and control the process of deduction in inquiry, i.e., they are prescriptive for the use of language and thought in inquiry, I argue that logic may, and should, be seen as an instrument or as a way of proceeding (*modus procedendi*) in inquiry.

KEYWORDS: empirical interpretation of logical truths, ontological interpretation of logical truths, semantic interpretation of logical truths, the nature of logical truths

I. Introduction

According to E. Nagel,[1] there are three main interpretations of logical principles.[2] One interpretation holds that logical principles are necessary truths which are descriptive of the most general structure of everything both actual and possible; the second interpretation maintains that they contingent, although very reliable, empirical hypotheses, and the third interpretation takes them to be void of factual content and, thus, arbitrary specifications for the construction of symbolic systems. No doubt, these interpretations are based on some assumptions, more or less problematical. Very roughly, the first interpretation seems to assume that we have *a priori* knowledge about at least some facts, i.e., about at least part of the real structure of the world. The second interpretation assumes that all principles involved in inquiry are empirical generalizations, although some of them are not directly subject to experimental refutation. Finally, the third interpretation assumes that if a principle lacks factual content then it is arbitrary, even though it

[1] Ernest Nagel, "Logic without Ontology," in *Naturalism and the Human Spirit*, ed. Yervant H. Krikorian (New York: Columbia University Press, 1944), 211.

[2] The term 'logical principle' is sometimes understood as referring to certain logical truths or logical laws. In this paper, however, I take 'logical principle' and 'logical law' to be synonymous with 'logical truth.' Although there could be made certain distinctions among these terms, for the purposes of this paper, I will not focus upon them.

© LOGOS & EPISTEME, VII, 3 (2016): 249-272

has an identifiable function in inquiry. Due to the strong arguments against them, all these three presuppositions are, as I will argue below, if not false, at least very problematical. In this paper, by disentangling the lack of factual content from arbitrariness, I will argue for, what may be seen as, a certain version of the third interpretation, according to which logical principles are propositions made true by the meanings of certain terms – the so-called logical terms – from a definite *linguistic framework*.[3]

The rationalistic assumption of the first interpretation seems very problematic due to the strong arguments against the existence of synthetic *a priori* knowledge *about facts*. Moreover, from an empiricist perspective, the *validity* of synthetic propositions is always subject to empirical tests and even if it holds in n cases, there is no logical guarantee that it will hold also in the *n+1* case, no matter how large *n* is; it follows that no proposition which has factual content can be necessarily true. Hence, once the rationalist view of knowledge is forsaken, i.e., the idea that reason considered independently can offer knowledge about facts, as A. J. Ayer[4] emphasized, the empiricist philosopher has to account for the logical principles in one of the following ways: "he must say either that they are not necessary, in which case he must account for the universal conviction that they are; or he must say that they have no factual content, and then he must explain how a proposition which is empty of all factual content can be true and useful and surprising." In other words, the empiricist has to decide whether logical principles are about the world, and, thus, not necessary or if they are necessary, but not about the world. This amounts, I believe, to a decision between the second and the third interpretations which Nagel mentioned, with the necessary emendations.

Regarding the structure of this paper, I will proceed as follows: I will first put forward certain methodological remarks with respect to the evaluation of the proposed interpretations. Second, in sections two and three, I will briefly present and critically evaluate two recent arguments for the ontological interpretation of logical principles (proposed by G. Sher and T. Tahko). In the forth section I will critically analyze three main instantiations [J. St. Mill, Quine, P. Maddy] of the idea that logical principles are empirical hypotheses. In the fifth section, I will present and argue for the idea that logical principles are true in virtue of the meanings of the logical terms from a certain *linguistic framework,* adopted for certain purposes of inquiry, purposes which also justify them. I will end by defending the proposed interpretation of two objections.

[3] I use the expression 'linguistic framework' in Carnap's sense, namely, a system of expressions together with the rules that govern their use (see section IV. b.).

[4] Alfred Ayer, *Language, Truth and Logic* (London: Penguin Books Ltd., 1936/1990), 65.

According to the interpretation that I put forward, logical principles are simply true in virtue of the meanings of the logical terms. Although their truth is independent of the facts from the world, they are non-arbitrary statements which are regulative for the use of language and deduction in inquiry. More precisely, logical principles specify the use of certain words and statements in inquiry. Since these principles also have a prescriptive function for the use of language and deduction in inquiry, I argue that logic – as a system of logical principles – may, and should, be seen as a way of proceeding (*modus procedendi)* in inquiry.

The idea that logic is an instrument for proceeding in (scientific) inquiry, or a *modus scientiarum*, was famously held by Aristotle and many mediaeval philosophers (e.g. Albertus Magnus, Aquinas, Petrus Hispanus). However, they argued that logical principles are at the same time principles of being, which, implicitly at least, makes them embrace the first interpretation mentioned above. Therefore, although the interpretation of logical principles defended in this paper has some features in common with the Aristotelian view, according to which logic is an *Organon,* i.e., an instrument, it should not be entirely associated with it.

II. Methodological Remarks

I think that it is important to briefly describe here what kind of methods, if any, could, and should, be used in order to evaluate the interpretations of logical principles mentioned above. These remarks will be useful for the particular analysis conducted in the sections below.

First, if logical principles are ontological principles that govern everything that is or could be, how could we test such a hypothesis? Do we have epistemic access, in principle, to everything that is or could be? Does this supposition have empirical consequences which could be tested? As far as I can see, this idea could not be effectively disproved. Nevertheless, I do not consider that it is meaningless, in a wide use of the term 'meaning,' but simply that its presuppositions are not sustainable.[5] On the one hand, it assumes that *reality* has such principles, and, on

[5] I think that what could be done when we confront ontological interpretations of logical principles – and this is the method that I will follow in this paper – is to criticize their presuppositions, and to show that such interpretations are not necessary for understanding the nature of logical principles and their role in inquiry. This idea was in fact explicitly stated by Ernest Nagel, who emphasized that "if philosophers propose to supply a foundation for logical principles by reading them as formulations of immutable and necessary structures of everything that is or could be, I know of no method for proving them in error. I believe nevertheless, that it is possible to dispense with such interpretations without impairing our understanding of the nature and power of logic." See Ernest Nagel, "In Defence of Logic without Metaphysics," *The Philosophical Review* 58 (1949): 34.

the other hand, it assumes that we are able to know them in an *a priori* manner. Hence, generally speaking, this interpretation maintains that we have *a priori* knowledge about certain relevant facts, although it indicates no ground for this assertion.[6]

Secondly, if logical principles are empirical generalizations, then they should be capable of being tested like all the other empirical hypotheses. However, as we will see in section IV of this paper, this criterion is not met by the logical principles.

Finally, if logical principles are true propositions in virtue of the meanings of the logical terms from a certain linguistic framework, we should be able to show that once we know the meanings of those terms, nothing else is required for establishing their truth. Moreover, once we have abandoned the idea that logical principles are grounded by the real structure of the world, which is supposed to guarantee their non-arbitrariness, we must explain why logical principles are non-arbitrary even in the absence of such a powerful link with reality.

III. Logical Principles as Ontological/Metaphysical Principles

The idea that logical principles are necessary principles of being has a longstanding tradition, and was famously supported by Aristotle. The principle of non-contradiction, one well-known and important logical principle, which is "the most certain of all principles" (*Metaphysics 1005b22*), is asserted by Aristotle, due to his general conception, as being true about facts: the same attribute cannot at the same time belong and not belong to the same subject in the same respect.

In the same spirit, Bertrand Russell also believed that "logic is concerned with the real world just as truly as zoology, though with its more abstract and general features."[7] It is very probable, however, that by this idea Russell was referring to the fact that abstract objects (like propositional functions), which are the subject matter of logic, are also part of the real world, and in this sense logic is also concerned with the real world.[8] The Swiss mathematician Ferdinand Gonseth, however, gave a nice expression of the idea that logic is concerned with the real

[6] The main problem with a view that asserts the existence of rational insights, as Boghossian puts it, is that "no-one has been able to explain, clearly enough, in what an act of rational insight could intelligibly consist." See Paul Boghossian, "Blind Reasoning," *Aristotelian Society Supplementary* 77 (2003): 230-231.

[7] Bertarnd Russell, *Introduction to Mathematical Philosophy* (London: George Allen & Unwin, Ltd., 1920, 2nd edition), 169.

[8] See Penelope Maddy, "The Philosophy of Logic," *The Bulletin of Symbolic Logic* 18 (2012): 497.

world, by saying that "logic is the physics of the arbitrary object,"[9] expression which also emphasizes the topic-neutral character of logic. Of course, whether we may have knowledge of such objects is a very problematical issue.

Even today, the idea that logical principles are primarily ontological principles is endorsed by some philosophers. For instance, T. Tahko expresses the principle of non-contradiction in a very similar manner as Aristotle did: the same attribute cannot at the same time belong and not belong to the same subject in the same respect and in the same domain.[10] In what follows I will briefly present and critically analyse two recent arguments, proposed by T. Tahko and G. Sher, for the idea that logical principles describe, or have a strong connection with, ontological/metaphysical structures.

a) T. Tahko's Metaphysical Interpretation of Logical Principles

Tahko's general idea is that logic is grounded in metaphysics, logical principles being supposed to express the most general structure of reality. Specifically, "a sentence is logically true if and only if it is true in every genuinely possible configuration of the world."[11] Thus, logical necessities might be explained as those propositions true in virtue of the nature of every situation, or every object and property. In addition, as he emphasizes, since only metaphysical modality could secure the correspondence between a possible world and the structure of reality, genuine possibility should be understood in terms of metaphysical possibility, preserving thus the idea that logic is the most general science. Metaphysics "is about mapping the fundamental structure of reality" and logic "is about representing the results formally."[12] Of course, since it is not necessary to formally represent the results of metaphysics, an immediate consequence of the latter idea is that logic would not be necessary for metaphysics, a view which is very implausible.

The metaphysical account for logical principles proposed by Tahko seems very problematic to me. In what way metaphysics maps "the fundamental structure of reality," and how exactly do we get to know, if it is possible, this fundamental structure of reality? If we suppose that this structure is to be known *a*

[9] Ferdinand Gonseth, *Qu'est-ce que la logique?* (Paris: Hermann, 1937).

[10] Tuomas E. Tahko, "The Metaphysical Interpretation of Logical Truth," in *The Metaphysics of Logic: Logical Realism, Logical Anti-Realism and All Things in Between*, ed. Penelope Rush (CUP, 2014), 239.

[11] Tahko, "The Metaphysical Interpretation," 239.

[12]Tuomas E. Tahko, "The Metaphysical Status of Logic," in *The Logica Yearbook 2007*, ed. Michal Peliš (Praha, Filosofia, 2008), 8.

posteriori, then we have no ground to say that it is the fundamental structure of reality, because experience offers us just contingent facts.[13] If we suppose that this structure is to be known *a priori*, as the metaphysicians usually believed, we come back to rationalism, but, as we mentioned above, also in this case we have no ground to assert that we have *a priori* knowledge about certain real facts.

In addition, as Nagel[14] similarly pointed out, when we say that logical principles are true in all genuinely possible configurations of the world (GPW), what do we mean by a 'genuinely possible configuration of the world?' If we identify a GPW on the basis of logical principles, namely, a GPW is a configuration of the world which conforms to logical principles, and there is no other way to identify a GPW, then we simply have a nominal, trivial definition. Namely, a GPW is a possible world which conforms to logical principles and thus they hold in each GPW. This definition simply gives the meaning of the expression 'GPW,' and there is no way in which such a definition may by refuted by any possible observations. However, in this case the definition of logical truths becomes circular, because the expression 'logical truth' also occurs in the *definiens*, namely: a sentence is a logical truth if and only if it is true in every world which conforms to logical truths. Of course, if a GPW is identified by metaphysical criteria, then we have the difficulties mentioned above.

Moreover, in the formulation of the principle of non-contradiction mentioned above, a very important role is played by the expressions 'same attribute' and 'same respect.' These specifications seem to be meant to save the principle for all counterexamples and, thus, make us unable to construct a genuine empirical test. The main idea is that the principle is employed as a criterion for specifying 'the same attribute' and 'the same respect.' Thus, the principle has a *self-protective formulation*. For example, if we take a coin and say that it is circular and also not-circular, it will be objected that not in the same respect (once viewed perpendicular to its faces, and then from the middle, parallel to its faces). If we specify the same respect as being the face of the coin viewed perpendicularly, the coin will delimit an angle of thirty degrees and also one of sixty degrees. In this case, the defender of the principle will say: yes, but not in the same respect; it is not viewed at the same distance from the face of the coin. In order to save the principle, what has been previously established as *the same respect* is now modified, i.e., the conditions in which we evaluate the previously

[13] This is in fact one of the main ideas of Wittgenstein's *Tractatus*, i.e., the view that we may have knowledge, in the precise sense of this term, only about contingent facts, and was also famously stated by David Hume. See also Ayer's reasoning from the *Introduction* section above.
[14] Nagel, "Logic without Ontology," 214-217.

established *same attribute* are now modified, and the principle of non-contradictions functions as a criterion for specifying the new 'the same respect.' We do not have a specification of 'the same respect' antecedent to the application of this principle. Thus, because of the way in which 'the same respect' is used, we cannot properly test the principle. More generally, since the expression 'the same respect' seems to belong to the epistemological lexicon and it is introduced in an ontological definition of the principle, the validity of this interpretation raises serious doubts.

Furthermore, if we consider the diameter of the coin and say that it has 2 centimeters, and then that it has 3 centimeters, it will be argued that it is not possible. But the impossibility does not come from empirical tests. The impossibility for the same diameter to have two dimensions, in the same time, derives from the fact that we use the expressions '2 centimeters' and '3 centimeters' to formulate different outcomes of measurement. No diameter will have two dimensions in the same time because the expressions are used in such a way that one of the attribute of dimension is used to specify the absence of the other. Hence, the underlying idea is that the 'sameness' and 'difference' of attributes are specified in terms of the conformity of attributes to the principle of non-contradiction. We have to apply the principle in specifying 'the same attribute' before deciding whether a certain controversial instance obeys or nor the principle of non-contradiction. This suggests that the principle of non-contradiction works as an instrument of specifying the use of expressions in a language, as a regulative principle for operating distinctions, rather than being an ontological principle.[15]

Finally, it worth mentioning that even the etymology of the word 'contra-diction' comes against an ontological explanation of the principle of non-contradiction. The Latin word 'contradictio' derives from 'contradico' which means 'speak against.' Thus, only a *dictum* can come against another *dictum*, but not an object, a fact or an event. In the spirit of this line of thought, David Hilbert emphasized in his lecture "On the Infinite" that to think that facts could contradict one another is simply 'careless thinking':

> As some people see ghosts, another writer seems to see contradictions even where no statements whatsoever have been made, viz., in the concrete world of sensation, the 'consistent functioning' of which he takes as special assumption. I myself have always supposed that only statements, and hypothesis insofar as they lead through deduction to statements, could contradict one another. The view

[15] For a similar discussion see also Nagel, "Logic without Ontology," 212-214, and Nagel, "In Defence of Logic," 29-30.

> that facts and events could themselves be in contradiction seems to me to be a prime example of careless thinking.[16]

Of course, a fellow of the ontological approach to the logical principles will easily accept that objects and events cannot, as a matter of fact, contradict one another, and this is precisely because the law of non-contradiction does not allow them. What Hilbert says, however, is more than that: he says that the facts or events could not contradict one another because the notion of contradiction cannot be meaningfully applied in the world of facts. That is to say that it makes no sense to assert that facts could or could not contradict one another. To apply the notion of contradiction in the domain of facts is simply a categorical error, an example of 'careless thinking.'

b) Gila Sher's Invariantist Interpretation

According to Gila Sher[17] logic "is grounded both in the mind and in the world, and its two grounds are interconnected." What Sher precisely understands by 'world' is not so clear, but, nevertheless, she clearly specifies that the terms 'world' and 'reality' (taken as synonyms) are not used to denote 'thing in itself,' 'mere appearances,' neither just empirical experience, not conceptual reality. In spite of these negative determinations, however, "logic is both in the mind and in the world in a substantive sense, a sense that yields significant explanations, solves significant problems, and has significant consequences."[18] Although this account is not a purely ontological one, the main features of this interpretation, as we will see below, endorse I believe the idea that Sher's account of logic is strongly related to an ontological interpretation of logical principles.

The main argument for this view regards the intimate relation between logic and reality *via* truth. The relation of logical consequence establishes between a set of sentences Γ and a sentence S if and only if the truth of Γ is transmitted to S, or *guarantees* the truth of S. However, since truth "inherently depends on whether things in the world are as given sentences say they are,"[19] then the notion of logical consequence also depends on the facts of the world. Specifically, in nontrivial cases, S is a logical consequence of Γ if the facts described by Γ *strongly*

[16] David Hilbert, "On the Infinite," translated by Erna Putnam and Gerald J. Massey from *Mathematische Annalen*, vol. 95, (Berlin, 1926), in *Philosophy of Mathematics: Selected Readings*, 2nd edition, ed. Paul Benacerraf and Hilary Putnam, (Cambridge University Press, 1983), 185.

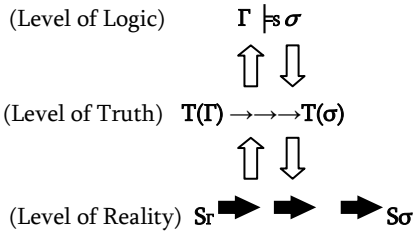
[17] Gila Sher, "Is Logic in the Mind or in the World?" *Synthese* 181 (2011): 354.

[18] Sher, "Is Logic in the Mind," 354.

[19] Sher, "Is Logic in the Mind," 356.

necessitate the facts described by S. More precisely, the main idea is that the relation of logical consequence is grounded by a formal *strong necessitation* relation present in reality, which establishes between states of affaires. This relation is a formal mathematical relation that governs "the formal (structural) features of objects, or their formal behaviour."[20] The notion of formality is defined in mathematical terms, by generalizing Tarski's criterion of logicality, namely, "to be formal is to be invariant under the isomorphisms of structures."[21]

Among the three relations just described (i.e., logical consequence, guarantee, and strong necessitation), there exist downward and upward dependencies, which are meant to ground the relation of logical consequence in reality. The downward dependency indicates that if the relation of strong necessitation does not obtain between the relevant states of affairs then neither the relation of guarantee, nor the relation of logical consequence, obtains. The upward dependency indicates that if certain premises logically imply a certain conclusion then the relation of strong necessitation obtains between the relevant states of affairs, namely, those described by the premises and conclusion. We may represent all these relations – as Sher[22] does – by different kind of arrows in the following diagram:



Although Sher's interpretation of logical consequence is very interesting, because it goes beyond the limits of possible experience,[23] it is open to criticism.

[20] Sher, "Is Logic in the Mind," 361-362.

[21] Sher, "Is Logic in the Mind," 363. See also Alfred Tarski, "What are Logical Notions?" *History and Philosophy of Logic* 7, 2 (1986): 143-154.

[22] Sher, "Is Logic in the Mind," 362.

[23] It is beyond the limits of possible experience because there are an infinite number of instances of logical implication, and we cannot verify whether all of them are grounded in something present in reality; we also lack a proof which shows that in principle they could be grounded in reality). In addition, we have no reason to assert that we have access to the real structure of reality, be it mathematical or not.

First, as Rossberg[24] indicates, there is no requirement to find actual situations in the world in order to show that the premises of an argument are true while the conclusion is false; any counter-model will do this job. Thus, a failure of the relation of strong necessitation seems unnecessary for grounding the failure of logical consequence. In addition, since classical logic is grounded in the worldly strong necessitation relations formulated by classical mathematics, and "in the case of nonclassical logic, the formal laws are given by nonclassical mathematics,"[25] we may wonder, as Rossberg[26] does, how is it possible that classical mathematics allows us to ground classical logic in reality, and intuitionist mathematics allows us to ground intuitionist logic in reality, and, yet, they disagree? For this may suggest that, after all, logic is not grounded in reality, but in the (mathematical) representation of reality. As a matter of fact, it would be a more modest assumption to suppose that mathematics "imposes structure on reality" rather than discovering the structure of reality, in which case "we have considerable freedom in the choice of structures that we want to give the world."[27]

In fact, even if we assume Sher's definition of formality, in order to fulfil its task, we must make explicit a necessary requirement for the mathematical theory which is meant to represent the structure of reality, namely, that it has to be categorical.[28] Thus, logical consequence could be grounded only in worldly formal relations represented by categorical mathematical theories. Moreover, of course, the proposed interpretation of the ground of logic assumes that we could know the real structure of reality. Still, since we are supposed to know this structure *via* mathematics, which is generally believed to be an *a priori* inquiry, then it also assumes an *a priori* knowledge about facts, i.e., about at least part of the real structure of the world. Furthermore, as a final remark, I think that Sher's interpretation only seems plausible because, as her particular examples illustrate,[29] it uses a set-theoretic interpretation of logical operators. Of course, this would not entail that logic is grounded in reality, but merely that we may interpret logical operators in set-theoretic terms.

---

[24] Marcus Rossberg, "Comment on Gila Sher's 'Is Logic in the Mind or in the World?'" Pacific APA, Vancouver, April 8-12 (2009): 3. Online version: http://homepages.uconn.edu/~mar08022/papers/Rossberg_on_Sher.pdf

[25] Sher, "Is Logic in the Mind," 364.

[26] Rossberg, "Comment," 9.

[27] Rossberg, "Comment," 9. The existence of different geometries may illustrate better this point with respect to *the* structure of space.

[28] It is well known that not all mathematical theories meet this criterion.

[29] For instance, the existential quantifier is interpreted as non-emptiness, conjunction as intersection, and so on.

To sum up, the idea that logical principles describe the most general structure of reality, or that they are grounded in such a structure, does not seem to be sustainable. First, since logical principles are taken in general to be known *a priori*, i.e., their truth is independent of observations, and also to describe at least some facts, i.e., real structures, the present interpretation assumes an *a priori* knowledge about facts. However, as we repeatedly emphasized, there is no reasonable ground for asserting this idea; we do not have knowledge of undetermined objects, of *objects as such*. Second, it seems to transform the function of logical principles for introducing distinctions and instituting adequate linguistic usage, into ontological constraints. Although it seems very plausible to interpret some logical principles in an ontological manner (at least for the level of the world accessible to our experience), we have no reasonable ground to maintain this. Therefore, this interpretation does not seem feasible; a better candidate that has less problematical assumptions would be preferable.

IV. Logical Principles as Empirical Generalizations

In this section I will critically analyze three main instantiations (Mill, Quine, Maddy) of the idea that logical principles are empirical hypotheses, and, thus not necessary. Maddy's interpretation, as we will see, although is an empirical one, takes them to be necessary only relative to the presence of the corresponding structure of the world – a view which needs some ontological underpinnings.

a) J. St. Mill's View

One of the pioneers who endorsed the idea that logical principles are not necessary propositions was J. St. Mill. For him, they are *a posteriori* and thus unnecessary. Mill believed that logical principles are inductive generalizations[30] confirmed in an extremely large number of cases. This large number of instances makes us to believe that logical principles are necessarily and universally true and

[30] Mill believed that principles such as the principle of non-contradiction, or of excluded middle, are real propositions, i.e., they convey new information, and not merely verbal, i.e., "which assert of a thing under a particular name only what is asserted of it in the fact of calling it by that name." John St. Mill, *A System of Logic Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation* (London: Longmans, Green and Co. 1886), 74/ Book I, Chap. VI. Being real, however, these propositions are, as for Quine, *a posteriori*. The ground for Mill's distinction between real and verbal propositions is to be found in his (semantic) theory of denotation and connotation (see John Skorupski, "Mill on Language and Logic," in *The Cambridge Companion to Mill*, ed. John Skorupski (Cambridge University Press, 1998), 36-40.)

that, although is possible, a negative instance will never appear. According to this view, the method for testing the validity of logical principles is the same as for the other empirical hypotheses, specifically, if an argument gives a materially true conclusion from materially true premises then it is valid, if not, it is invalid.[31] Consequently, in order to establish the validity of an argument we need empirical evidence.

We may agree, however, that logical principles could be discovered and learned inductively, but this does not entail that they are known, or could only be known, empirically. As we will argue below, logical principles may be known independently of experience. By this I mean, following Ayer,[32] that their validity is not determined in the same way as for the empirical hypotheses. For instance, let us consider an argument from whose premises 'A' and "if A then B," asserted as true, is drawn – according to the rule modus ponens[33] – the conclusion 'B,' which, as a matter of fact, is false.[34] If we follow the proposed method, then we will have to reject *modus ponens* as a universally valid rule. But it seems that in such a case, as long as the normal meanings of the logical terms are preserved, we are more inclined to say that the premises were asserted mistakenly or that the recognition that 'B' is false was an error. There is no doubt that the proposition "If A and (if A then B), then B" is true as long as the terms 'and' and 'if… then' have the meanings as given by the normal truth tables.[35]

Moreover, we know that the validity of many hypotheses employed in science can only be established by examining the consequences implied by them in accordance with logical principles. Nevertheless, in a non-holistic context,

[31] This particular method seems to be implicitly present also in Sher's account, because she believes that if a certain relation does not establish between the states of affairs represented by the sentences of an argument, then the argument is invalid, i.e., the relation of logical consequence does not establish either. In Sher's terms, a failure of strong necessitation relation entails the failure of the corresponding relation of logical consequence.

[32] Ayer, *Language, Truth and Logic,* 68.

[33] Mill had in mind the Aristotelian logic, but his considerations may be applied also to modern logic.

[34] Such interpretations, supposed to be counterexamples to modus ponens, were in fact proposed by Vann McGee, "A Counterexample to Modus Ponens," *The Journal of Philosophy* 82 (1985): 462-471 and Niko Kolodny and John MacFarlane, "Ifs and Oughts," *Journal of Philosophy* 107(2010): 115-143, and they have generated ample discussions among logicians and philosophers.

[35] See also Nagel, "Logic without Ontology," 219 and Constantin C. Brîncuş and Iulian D. Toader, "A Carnapian Approach to Counterexamples to Modus Ponens," *Romanian Journal of Analytic Philosophy* VII (2013): 78-85.

when the consequences derived from premises believed to be true are in disagreement with the observations of experience, it is typically not the logical principles used to drawn the consequences which are rejected. If they where, then the relation of logical consequence would be an empirical one, and it would be difficult to speak about the confirmation or confutation of hypothesis by empirical data. It follows that the proposed method for testing the logical principles is not a feasible one. As long as we accept that we can test certain domains of science singularly, i.e., we disprove the holistic view, we should accept the idea that the ground for the revision of logical principles must lie elsewhere than in the subject matter of the natural sciences – in the sense that observations could not directly refute a logical principle. In the next section I will argue that the situation is the same even in a holistic context.

b) Quine's Naturalist[36] Approach

A more sophisticated form of empiricism was elaborated by W.V.O. Quine, who embraces the first option that the empiricist, according to Ayer, has available, namely, logical principles are about the world, and, thus, non-necessary. According to Quine, since "logic, as any science, has as its business the pursuit of truth"[37] and "there is no higher access to truth than empirically testable hypotheses,"[38] it follows that logic, as the entire human knowledge, has the same status, namely, it is *a posteriori*. Logical principles are themselves a constituent part of *the entire system of science*, and, consequently, they also confront, although indirectly, the experience tribunal. Indirectly because, according to Quine, what we actually test are not isolated propositions, or particular sets of propositions, but the entire system of science. In the case of a conflict with experience we may revise, in accordance with the principles of conservatism and simplicity, whatever proposition from the system.[39]

[36] Quine's conception on the nature of logical principles does not necessarily follow from his holistic view – Carnap himself adopts the epistemological holism, but mainly from his attack of the first 'dogma' of empiricism, which leads finally to the naturalistic representation of knowledge, i.e., to the idea that all our knowledge is a posteriori. Epistemological holism and revisability of any statement are perfectly compatible with the existence of a clear and precise distinction between a priori and empirical knowledge (see Michael Friedman, "Philosophical Naturalism," *Proceedings and Addresses of the American Philosophical Association* 71(1997): 9-10.)

[37] W.V.O. Quine, *Methods of Logic (revised edition)* (New York: Holt, Rinehart and Winston, 1950/1966), xi.

[38] W.V.O. Quine, "Naturalism; Or, Living within One's Means," *Dialectica* 49 (1995): 251.

[39] See W.V.O. Quine, "Two Dogmas of Empiricism," *Philosophical Review* 60 (1951): 20-43.

It is important to emphasize, however, that Quine does not endorse the idea that we establish the validity of logical principles by confronting them with observational data, in order to see if materially true premises entail a materially true conclusion. The revision of a logical principle is made as a pragmatic decision for readjusting the entire system of science to observational data. Logical principles can be revised, "but this is not to deny that such laws are true in virtue of the conceptual scheme, or by virtue of meanings," and "because these laws are so central, any revision of them is felt to be the adoption of a new conceptual scheme, the imposition of new meanings on old words."[40] This amounts, I believe, to saying that logical principles are true in virtue of the meanings of the logical terms, and to the recognition of the fact that the meanings of such terms could be changed.[41]

However, it seems to me that there is an important difference between the revisions of truth-values of empirical statements, whose meanings are preserved, and the revision of the truth-values of statements by changing their meanings it is an important difference. In my understanding, this entails the idea that there is a distinction between propositions true in virtue of meanings, and propositions true in virtue of facts, i.e., between analytic and synthetic propositions, even if such a distinction may admit borderline cases with respect to the entire system of science. In spite of this, the fact that logical principles are revisable does not entail that they are not necessary and, consequently, empirical generalizations. As we will see below, although they could be revised, logical principles are true independent of facts, and thus necessary, in a certain linguistic framework.

In some writings,[42] Quine seems to rule out any kind of distinction between analytic and synthetic propositions, suggesting that all sentences have, in a certain degree, empirical content, i.e., they all are synthetic. For instance, he believes that the validity of mathematics is established by confronting it with the observational data. This happens because when we test an empirical hypothesis we take it often in conjunction with propositions from pure mathematics. In this way pure mathematics becomes applied. If the theory is corroborated by experiments, then mathematical propositions are believed to be true, if not they are refuted.

[40] Quine, *Methods of Logic*, xiv.

[41] In *Philosophy of Logic*, (Harvard University Press, 1994), 81-82, Quine emphasizes that logical terms change their meanings in different logics. A change of logic amounts, thus, to a change of subject, i.e., a change of the meanings of the logical terms. In this respect, Quine is in agreement with M. Dummett who also considers that when two different logical schools disagree, they understand some logical terms in different ways. See Michael A. E. Dummett, *The Logical Basis of Metaphysics* (London: Duckworth, 1991), 302.

[42] Quine, "Naturalism," 251- 261.

However, as M. Friedman emphasized, the fundamental problem with this representation is that a physical theory, viz. the theory of relativity, is not happily viewed as a large conjunction formed from Einstein field equations, the Kleinian theory of transformation groups, and the Riemannian theory of manifolds, in which case Eddington's experimental results "are potentially spreading empirical confirmation over the entire conjunction."[43] In such cases the mathematical conjunct works rather "as a necessary presupposition of that theory, as a means of representation or a language, as it were, without which the theory could not even be formulated or envisioned as a possibility in the first place."[44] This amounts, in my understanding, to recognize the fact that there is a distinction between propositions from empirical science, i.e., synthetic, and analytic propositions which work as instruments in the system of science, and whose truth is not a problem of matter of facts, but of meanings.[45]

We can, and should, admit that logical principles are revisable, but, following Carnap, who otherwise agrees with many of Quine's ideas,[46] we should recognize a distinction between the revision of the truth-values of certain propositions on empirical grounds, without abrogating their meanings, and the

[43] Friedman, "Philosophical Naturalism,"12.

[44] Friedman, "Philosophical Naturalism,"12.

[45] Friedman's reply also answers Alonzo Church's objection to Nagel's idea that logical principles are not tested in the same manner as the empirical hypotheses (see Alonzo Church, "Review: Ernest Nagel, 'Logic without Ontology'," *The Journal of Symbolic Logic* 10 (1945): 17. Logical principles, and probably the mathematical ones, are not conjuncts in the entire system of science which confronts the experience tribunal, but rather they are regulative principles which also serve as conditions for formulating certain empirical hypotheses. The relation between logico-mathematical statements and the other statements is not that of conjunction but rather of presupposing, which is a very different relation. As N. Rescher puts it, "p presupposes q means 'q is a necessary condition for the very possibility (or even meaningfulness) of p'". Formally: ($\Diamond p \rightarrow q$). See Nicholas Rescher, "On the Logic of Presupposition," *Philosophy and Phenomenological Research* 21 (1961): 527.

[46] "Quine shows that a scientist, who discovers a conflict between his observations and his theory and who is therefore compelled to make a readjustment somewhere in the total system of science, has much latitude with respect to the place where a change is to be made. In this procedure, no statement is immune to revision, not even the statements of logic and of mathematics. There are only practical differences, and these are differences in degree, inasmuch as a scientist is usually less willing to abandon a previously accepted general empirical law than a single observation sentence, and still less willing to abandon a law of logic or of mathematics. With all this I am entirely in agreement." Rudolf Carnap, "W. V. Quine on Logical Truth," in *The Library of Living Philosophers, Vol. XI, The Philosophy of Rudolf Carnap*, ed. Paul Arthur Schilpp, (Open Court Publishing Company, 1963/1997), 921.

revision of the truth-values of certain propositions by changing their meanings. I think that Carnap's remarks[47] are helpful for understanding this distinction:

I should make a distinction between two kinds of readjustment in the case of a conflict with experience, namely, between a change in the language, and a mere change in or addition of, a truth-value ascribed to an indeterminate statement, (i.e., a statement whose truth value it not fixed by the rules of language, say by the postulates of logic, mathematics, and physics). A change of the first kind constitutes a radical alteration, sometimes a revolution, and it occurs only at certain historically decisive points in the development of science. On the other hand, changes of the second kind occur every minute. A change of the first kind constitutes, strictly speaking, a transition from a language L_n to a new language L_{n+1}. My concept of analyticity as an explicandum has nothing to do with such a transition. It refers in each case to just one language; 'analytic in L_n' and 'analytic in L_{n+1}' are two different concepts. That a certain sentence S is analytic in L_n means only something about the status of S within the language L_n; as has often been said, it means that the truth of S in L_n is based on the meanings in L_n of the terms occurring in S.

Whenever a change of the first kind occurs, such change is made as a pragmatic decision for readjusting the entire system of beliefs for certain purposes of inquiry. The decision of changing a linguistic framework, i.e., a system of expressions together with rules that govern their use, is not in itself a cognitive matter, although it may, nevertheless, be influenced by theoretical knowledge.[48] Therefore, logical principles, analytic[49] principles in a certain language, are true in virtue of the meanings of the logical terms from that language, and can be revised once we make the pragmatic decision to change it (see section V for the idea that logical principles are 'framework principles').

[47] Carnap, "W.V. Quine on Logical Truth," 921.

[48] See Rudolf Carnap, "Empiricism, Semantics, and Ontology," *Revue Internationale de Philosophie* 4(1950): 20-40.

[49] There is a distinction between statements true in virtue of the logical terms (logical truths) and statements true in virtue of logical and non-logical terms (analytic statements *per se*). However, if we define the analytic statements as statements true in virtue of meanings, then, in this sense, logical truths are also analytic. In this context of the discussion, the distinction is not so relevant.

c) Maddy's Second Philosophy Account

Another interesting view of logical principles was recently proposed by Penelope Maddy,[50] who develops an empirical interpretation starting from the Kantian combination between transcendental Idealism and empirical Realism. According to Kant, logical structure, viewed transcendentally, is imposed on the world by our discursive modes of thought, and, viewed empirically, the world simply displays those structures as a matter of objective fact. Maddy tries to preserve these two features in a naturalized framework, by arguing, for the empirical side first, that the macro-world simply displays a certain structure, a Kant-Frege (KF) structure (given by the Kantian forms of judgement and updated with the Fregean results, and formed from objects, properties, relations, dependencies), and then arguing, for the naturalized transcendental side, that our cognitive mechanisms have evolved in such way that are able to detect this KF structure. The logic which represents, or is true of, this KF structure, however, is not identical with the entire classical logic, because 'the physical structure of the world' does not validate all principles of classical logic. The law of excluded middle and the material conditional "appear as idealizations introduced into that logic for good reasons."[51]

In sum, Maddy's idealized inquirer, the Second Philosopher, believes that the macro-world really has a KF structure, and that our cognitive mechanisms detect this structure because we live in a KF world and interact with it. These ideas are sustained by a large number of recent psychological studies, i.e., experimental studies, which are meant to support the idea that we are able we detect objects, properties and relations because they are really there, in the world. In the sketched picture, "logical truths are true because the world is made up of objects enjoying various interrelations with dependencies between them, and we tend to believe some of the simpler of these truths because human cognition has been turned by evolution to detect these very features."[52] Nevertheless, since the structure observed in our experience seems not to be present, for example, at the (quantum) micro-world, then we must admit that "logic applies to a situation insofar as it does have those features, and our cognitive machinery has evolved to detect those features." Therefore, the updated definition becomes: "logical laws are

[50] Penelope Maddy, *Second Philosophy* (Oxford University Press, 2007); Penelope Maddy, "The Philosophy of Logic," *The Bulletin of Symbolic Logic* 18, (2012): 481-504.
[51] Maddy, "The Philosophy of Logic," 500.
[52] Maddy, "The Philosophy of Logic," 501.

true in any situation with the right physical structuring; their truth is contingent on the presence of that structuring."[53] Moreover, Maddy emphasizes that

> we tend to believe the laws of logic independently of any experience because of our hard-wiring, we know them in a sense *a priori*, and we tend to think of them as necessary, that is, we tend to built them into our very idea of a possible world – and all this happens despite the fact that they wouldn't be true if the world were different and in fact don't seem to hold in the actual micro-world.[54]

Although I find this proposal very interesting, I am very sceptic regarding its validity. Even if we may agree that we usually observe a so-called KF structure in the world that we live in, this does not necessarily entail that the (macro-) world really has this structure, i.e., that the KF structure is the real structure of the macro-world. I think that the psychological observations do not offer us a sufficient ground for inferring that the structure we observe is the real structure of the macro-world, i.e., of a certain level of the world. Since psychological studies are based on observations, that are always made in a 'horizon of expectations'[55] which, in turn, reflects the manner human beings approach the world, it follows that observations do not represent pure facts of the world, or its fundamental structure. They are always relative to the human point of view. Thus, although it starts as an empirical interpretation of logical principles, this account is transformed in a *relativized* ontological interpretation. 'Relativized' in the sense that considers the world to have certain different structures at different levels and, due to the fact that we live in a certain domain/at a certain level of the world, we have access to the very structure of (this level of) the world.

To sum up, the interpretation of logical principles as empirical hypotheses, which are true in virtue of empirical facts, is not feasible. Mill's vision seems untenable because it disregards certain logical facts, i.e., the way in which logicians test validity of logical propositions, and the way in which the method of science actually works, namely, it presupposes the validity of logical principles, in deriving consequences from general hypotheses, and is not aiming at validating them. Quine's vision is not essentially problematic because it is holistic, Carnap also accepts the epistemological holism, but because it seems to disregard the distinction between propositions true in virtue of meanings and propositions true in virtue of facts, and, consequently, the kinds of changes that may occur in the entire system of science. The recognition of this distinction means, implicitly, that

[53] Maddy, "The Philosophy of Logic," 502.

[54] Maddy, "The Philosophy of Logic," 502.

[55] See Karl R. Popper, "The Bucket and the Searchlight: Two Theories of Knowledge," in Karl R. Popper, *Objective Knowledge. An Evolutionary Approach* (OUP, 1979).

logical principles are true in virtue of the meanings of logical terms from a certain linguistic framework. This point will be elaborated in section five. Finally, Maddy's interpretation[56] seems to me to be closely related to an ontological one, by presupposing that we come to know the real structure of the world, and by implicitly assuming that these structures are *reflected* in an invariant way by language.

V. Logical Principles as Regulative Principles of Inquiry[57]

In general, natural language, as it is, is sufficient for the purposes of efficient communication in daily activities. However, in certain domains of inquiry, especially in science, a greater precision is necessary for the use of language than the one found in natural language. For instance, to take a trivial example, a certain term must express the same meaning in the context of an argument, and this is precisely what the principle of identity – in one of its formulations – requires. In the same manner, the principle of non-contradiction requires that a certain term should not be applied and denied to the same object in the context of an argument. People do not always follow the rule modus ponens in their ordinary reasoning, but this desideratum of logic must be followed in science. In this sense, logical principles have a prescriptive function for the use of language. They do not describe the actual way in which agents think and use language.[58] They indicate the direction in which precision may be obtained, and, therefore, they fix an ideal that may, and should, be achieved in order to fulfil certain objectives of inquiry.

Let us consider for instance the various modern systems of logic. Their main aim is not to represent the 'true nature,' if any, of an antecedently identifiable relation of 'implication;' they are built as alternative specifications for a precise use of this term and for the performance of inferences.[59] Without explicit logical

[56] An interpretation that takes logical principles as a product of evolution, without assuming that they have, or are grounded by, a corresponding structure, would be less problematical. They could be seen is as instruments adopted in the course of evolution for their adaptability function, which also justifies them.

[57] This section develops, and is mainly based on, Rudolf Carnap and Ernest Nagel's interpretations of logical principles and on the interpretation developed by (other) logical positivists (viz. A.J. Ayer, H. Hahn et al.).

[58] The psychologistic conception, which states that logical truths are empirical statements which describe the ways in which people actually think, has been in a continuous obliteration after Gottlob Frege's well-known criticisms, according to which logic is concerned with the ways in which people *must* think, if they are not to miss the truth.

[59] As a matter of fact, Quine himself regards the theory of deduction (for propositional logic) as "a formal systematization of certain aspects of the ordinary use of language and exercise of

principles it is almost impossible to evaluate the validity of the performed inferences. Once the meanings of certain terms – the so-called logical terms – are precisely fixed, inferences can be performed and evaluated in a precise manner. Moreover, the fact that the meanings of logical terms from a system of logic do not correspond to the meanings of their counter-parts from natural language show us why logical principles also serve as "proposals for modifying old usages and instituting new ones"[60] and, thus, their regulative function is again revealed. Their main aim is to direct the use of language in the direction of clarity and precision.

The idea that logical principles are true in virtue of the meanings of the logical terms, we may say, is obvious from the practice of logic. In order to see that a statement is a logical truth, we do not make appeal to any facts, we simply apply the semantic and syntactic methods which are essentially based on the meanings of the logical terms – no matter how we may take these meanings to be defined, via model-theory or via proof-theory (as the inferentialists do). It is important to emphasize the difference between the idea that logical truths are based on linguistic conventions,[61] and the idea that they are true based on meanings. Rudolf Carnap himself disapproved the expression "linguistic conventions" as applying to his explanation of logical truths. The choice of the meanings of the logical terms may be a matter of convention, but once these meanings are fixed, there is not conventional at all which statements are logically true: "once the meanings of the individual words in a sentence… are given (which may be regarded as a matter of convention), then it is no longer a matter of convention or of arbitrary choice whether or not to regard the sentence as true; the truth of such sentence is determined by the logical relations holding between given meanings."[62]

Logical principles are also necessary relative to the meanings we attribute to the logical terms. If we change those meanings, then we must hold a different

---

reason." See W.V.O. Quine, "Ontological Remarks on the Propositional Calculus," *Mind, New Series* 43 (1934): 473.

[60] Nagel, "Logic without Ontology," 227.

[61] I do not endorse the idea, as Nagel, in "Logic without Ontology," does, that logical truths are linguistic conventions or consequences of such conventions, given, probably, by implicit definitions. In this way, Quine's famous criticism for the "linguistic theory of logical truth," a label given by Quine, may be putted aside. In fact, as I mentioned, Carnap found this description inappropriate for his explanation of logical truth. Azzouni's recent article on logical conventionalism offers a good analysis of Quine's criticism of logical conventionalism. See Jody Azzouni, "A Defense of Logical Conventionalism," in *The Metaphysics of Logic: Logical Realism, Logical Anti-Realism and All Things in Between*, ed. Penelope Rush (CUP, 2014).

[62] Carnap, "W. V. Quine on Logical Truth," 915-916.

class of logical principles. Of course, there is nothing necessary in maintaining a certain class of meanings for certain words. The fact that a certain choice of meanings was fruitful in the past does not guarantee that it will be fruitful in the future. Nevertheless, the truth of certain logical principles, once certain meanings for logical terms were established, is different from the acceptance of those meanings in future. The acceptance of those meanings is a pragmatic decision which, once accepted, entails a certain class of logical principles.

The idea that logical truths are true in virtue of meanings, i.e., analytic, necessary and prescriptive is fruitfully explained by Carnap in his article "Empiricism, Semantics, and Ontology," with the help of the concept of linguistic framework. As analytic statements, logical principles describe a linguistic framework. They are constitutive for a certain framework by providing the grammar and the rules for operating in that framework. In this sense they are necessary precisely because they are constitutive for the framework. Once you disobey them, you simply refuse to work within that framework. It is analogous with playing a game. If you do not accept the rules of a game, then you do not play *that* game. For that game, for that linguistic framework, the rules are constitutive, and thus necessary – from this internal perspective. The framework, of course, on pragmatic reasons, may be changed; its adoption is a contingent matter. This characterization of logical truths, as 'framework principles,' also reveals their regulative function. Since they indicate how one should work in a given framework, they are regulative for the activities performed in that framework.

Although the regulative function of logical principles is usually recognized, the objection often raised is that in order to formulate a reasonable ideal, and not an arbitrary one, logical principles must have an objective ground, namely, a ground, or a corresponding structure in reality. We may admire, however, this lofty rationalist ideal to ground logical principles in the structure of reality, but we are by no means forced to infer the arbitrariness of logical principles from the fact that they do not have an identifiable correspondent in reality. Human communication and inquiry are directed to the achievement of certain purposes, and it is a matter of fact that the objectives of communication and inquiry are better achieved when the language is used in the manner prescribed by logical principles. An empirical study of the behaviour of men employed in communication and inquiry confirms this idea. Therefore, even though logical principles do not have a ground, or a subject matter, in reality, this does not imply that they are arbitrary. The general idea mentioned above is that the justification of logical principles is better understood in terms of *objectives to be attained*.

More specifically, a set of logical principles is justified, if it is adequate for attaining certain purposes in inquiry. In this sense, the selection of a set of logical principles, instead of another, has an objective basis.

To sum up, logical principles are true statements in virtue of the meanings of the logical terms from a certain linguistic framework, in Carnap's sense discussed in section IV.b. To understand them is sufficient for determining their truth value. These principles, as long as the relevant meanings are preserved, are necessary because to deny them merely means to misunderstand the expressions from their structure (see the answer to the second objection from the next section).

VI. Final Remarks

The main aim of this paper was to present and to briefly analyze the main interpretations of logical principles. I have first presented the central features of the ontological (or metaphysical) interpretation of logical principles (Tahko, Sher), which was found infeasible because, in my understanding, it assumes, without a reasonable ground, an *a priori* knowledge about certain facts, and also seems, at least in Tahko's case, circular. Second, I have analyzed three main instantiations of the idea that logical principles are empirical hypotheses (Mill, Quine, Maddy), and I have tried to show why they seem problematic. Finally, I have sketched the main features of an interpretation which considers logical principles as non-arbitrary statements, regulative for the use of language in inquiry, in the direction of clarity and precision. According to this interpretation logical principles are true statements based on the meanings of the logical terms from a certain linguistic framework. Logical principles are necessary relative to the preservation of those meanings. The pragmatic decision to change the linguistic framework may entail the adoption of another set of logical principles, but, of course, this does not mean that logical principles are refuted by facts (as we argued in section IV). I will end now by considering two objections for the interpretation proposed in this paper.

An objection recently raised by Maddy[63] to the idea that logical truths are true only in virtue of the meanings of the logical terms, and that their truth does not depend (also) on facts from the world/our experience, is that our use of language is not independent of the facts from the world we live in, which shape our use of language. This would entail that logical truths are also dependent of some relevant facts. Therefore, the question from the title of this paper – *what*

[63] Maddy, "The Philosophy of Logic," 490.

makes logical truths true? – would not get its entire answer by pointing out only to language, or meanings.

I think that this objection could be dismissed. Of course, we may agree that natural language has an historical development and that the meanings of certain words *may be suggested* by our experience from the world that we live in, but this is not relevant for answering the proposed question, i.e., what makes logical truth true?. For instance, we may either follow Einstein[64] in saying that all our concepts and linguistic expressions – viewed logically – are free creations of our mind and could not be abstracted from experience, or we may agree that some concepts might be, somehow, suggested by experience, but this would not change the fact that the relevant factors for determining the truth of logical principles are only the meanings of the logical terms.[65] The issue raised by Maddy is relevant, I think, only for the problem of the origin of meanings, but since is sufficient to fully understand the meanings of the logical terms in order to establish the truth value of a logical sentence, the semantic conception of logical truth remains untouched.

Another objection often raised to the interpretation of logical principles as analytic statements, i.e., true in virtue of meanings, is that this view leaves unexplained the usefulness of logic in epistemic contexts, especially in the growth process of knowledge. I think that this is not the case. For instance, since the truth of logical principles is grounded in the meanings of the logical terms, we may ask ourselves: why these terms are introduced into language? As Hans Hahn[66] emphasized, a very plausible reason seems to be that we are not omniscient. Logical principles and logical deductions have significance for us precisely because we are not omniscient. If we were omniscient, then we probably would make only categorical assertions, without using logical terms as 'not' or 'or.' To use Hahn's example, if I am asked about the colour of the dress worn by Miss Erna yesterday, and I am not able to remember its colour, I could say: it was red or blue, or it was not yellow, but if I were omniscient, I would simply say: it was red (involving in this way no logical term).

Logical inference makes us aware of the propositions implicitly asserted when we assert other propositions – and it is in virtue of this fact that valid

---

[64] Albert Einstein, "Remarks on Bertrand Russell's Theory of Knowledge," in *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schilpp (New York: Tudor Pub. Co., 1952).

[65] The knowledge of the syntactic structure is, of course, presupposed in this context.

[66] Hans Hahn, "Logic, Mathematics, and Knowledge of Nature," in *Logical Positivism*, ed. A.J. Ayer (New York: The Free Press, 1959), 157.

inferences have epistemic significance.[67] For instance, if I assert that object A is either red or blue, and I also assert that object A is not red, then I implicitly have asserted that object A is blue. In this case, the conclusion is derived only in virtue of our rules which govern the use of the words 'or' and 'not,' and is not based on real connections among states of affairs, which we apprehend in thought. If someone refuses to recognize this valid logical deduction, he/she would not manifest a different belief about the behaviour of things, but he/she would merely refuse to speak about things according to the same rules as most of us do.[68] As long as we maintain certain rules for the use of expressions, we preserve the meanings of logical terms, and, thus, logical principles cannot be false; any denial of them would be self-contradictory – at least as long as the classical meaning of negation remains invariant. This is precisely why logical principles are necessary in a certain linguistic framework.[69]

[67] See Constantin C. Brîncuș, "The Epistemic Significance of Valid Inference – A Model-Theoretic Approach," in *Meaning and Truth*, ed. Sorin Costreie and Mircea Dumitru, (Bucharest: Pro Universitaria Publishing House, 2015), 11-36.

[68] See also Hahn, "Logic, Mathematics, and Knowledge of Nature," 156.

[69]**Acknowledgments**. The ideas from this paper were partially presented at the International Conference "Humanities and Social Sciences Today. Classical and Contemporary Issues," organized in the *Innovation and Development in the Patterning and Representation of Knowledge through PhD and Post-PhD Fellowships*, POSDRU/159/1.5/S/133675 Project, Romanian Academy, Iași Branch, 7-10 May 2015, and at the CELFIS Seminar on 25 November 2015 at the Faculty of Philosophy, University of Bucharest. I want to thank the audience, and especially Sorin Bangu, Sorin Costreie, Mircea Dumitru, Radu Ionicioiu, Dana Jalobeanu, Hamdi Mlika, Ilie Pârvu and Mihai Rusu for their useful remarks. I also want to thank Mircea Flonta and Iulian Toader for helpful comments and suggestions. Research for this paper was supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/159/1.5/S/133675.

RECOGNIZING 'TRUTH' IN CHINESE PHILOSOPHY

Lajos BRONS

ABSTRACT: The debate about truth in Chinese philosophy raises the methodological question How to recognize 'truth' in some non-Western tradition of thought? In case of Chinese philosophy it is commonly assumed that the dispute concerns a single question, but a distinction needs to be made between the property of *truth*, the concept of TRUTH, and the word ·truth·. The property of *truth* is what makes something true; the concept of TRUTH is our understanding of *truth*; and ·truth· is the word we use to express that understanding. Almost all human beings over the age of 2 have the concept of TRUTH, and therefore, the question whether some tradition has the concept of TRUTH is moot, but that doesn't imply that every language has a (single) word for ·truth·. Furthermore, recognizing ·truth· is complicated by the conceptual neighbors of TRUTH. What distinguishes ·truth· from its neighbors is disquotationality. Theories of *truth* similarly need to be distinguished from theories about adjacent notions. If a theory is more plausibly interpreted as a theory of *justification*, then it is not a theory of *truth*.

KEYWORDS: Chinese philosophy, comparative philosophy, concept of truth, theory of truth, truth

Introduction

Ever since Chad Hansen argued that (pre-Buddhist) "Chinese philosophy has no concept of truth,"[1] the role and nature of truth in ancient Chinese philosophy has been a hotly debate topic.[2] Much of this debate is plagued, however, by a confusion of terms, concepts, and theories of truth. Some of this confusion may be caused by Hansen's peculiar claim that "a concept is a role in a theory."[3] By that standard almost no one has a concept of truth, because – aside from a few philosophers – almost no one has a theory of truth. By that standard one may even doubt that Aristotle (and Plato) had a concept of truth as Aristotle's often quoted

[1] Chad Hansen, "Chinese Language, Chinese Philosophy, and 'Truth'," *Journal of Asian Studies* 44, 3 (1985): 492.

[2] For an introduction to and an overview of this debate, see: Alexus McLeod, *Theories of Truth in Chinese Philosophy* (London: Rowman & Littlefield, 2015), chapter 2. Alternatively, an easy way to get a list of contributions to this debate is to search for publications that refer to Hansen, "Chinese Language."

[3] Hansen, "Chinese Language," 504.

remark that "to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true" hardly qualifies as a *theory* of truth.[4]

The debate about concepts and theories of truth in ancient Chinese philosophy raises a methodological question: How does one recognize a concept and/or theory of truth in a tradition of thought other than Western philosophy? In the aforementioned debate, this question is rarely addressed, and it has received even less attention from scholars working on/with other traditions of (philosophical) thought. It is this question that this paper aims to answer.

In the introduction of his book on primitivism about truth, Jamin Asay points out that it is "absolutely vital" to distinguish the property of *truth*, the concept of TRUTH, and the word 'truth.' The first is "that feature (if it exists) that all truths share and all falsities lack;" the second is "our mental understanding of that notion that we use the word 'truth' to pick out;" and the third is, of course, 'truth' itself, but also 'true,' "is true," and so forth.[5] To keep the three apart, Asay writes the property as *truth*, the concept as TRUTH, and the word as 'truth,' and I will adopt this convention, with one minor adaptation, to be explained shortly.

The distinction is not specific to Western thought about truth, but applies to ancient Chinese philosophy as much as it does to any philosophy of truth. Insufficient attention to the distinction is not typical of the debate on Chinese philosophy either: many Western 'theories of truth' are about both property and concept, often confusing the two, and Anna Wierzbicka and associates' research on semantic primitives (see next section), for example, is about both concepts and words. Plural 'words' in the latter case, as Wierzbicka's research is about many other languages than English, languages that do not have the word 'truth,' but that have other words with the same functional role in those languages. Similarly, in case of other languages (than English), such as classical Chinese, we are not interested in the question whether it had the word 'truth' (because the answer to that question would obviously be "No"), but – borrowing Wilfrid Sellars's notational device of *dot quotation* – in whether that language included ·truth·.[6] A word or expression is ·truth· (or ·true·) in some language if it is playing the role in that language that is played by 'truth' (or 'true') in English.

---

[4] Aristotle, *Metaphysics*, in *The Complete Works of Aristotle*, ed. Jonathan Barnes (Princeton: Princeton University Press), 1011b25. Plato made similar remarks in *Cratylus* 385b2 and *Sophist* 263b. See also section 3.

[5] Jamin Asay, *The Primitivist Theory of Truth* (Cambridge: Cambridge University Press, 2013), 14.

[6] Wilfrid Sellars, "Abstract Entities," in *In the Space of Reasons*, ed. Kevin Scharp and Robert B. Brandom (Cambridge MA: Harvard University Press, 1963), 163-205.

By implication of the foregoing, the following three questions are different questions that need to be kept apart and answered separately:

1)Does *x* have a theory about the property of *truth*?

2)Does *x* have the concept of TRUTH?

3)Does *x*'s language include ·truth·?

In which *x* stands for 'ancient Chinese philosophy' or 'Polynesian philosophy' or any other non-Western tradition of (philosophical) thought that is the object of attention.

Nevertheless, these questions are not independent from each other. Arguably, one cannot have ·truth· without TRUTH (but there is no reason why the reverse would be impossible), and neither can one have a theory about the property of *truth* without TRUTH and ·truth·. Hence, of these three questions, the second is the most fundamental. For that reason, I will discuss how to answer that question first, before turning to the third and first (in that order). It needs to be emphasized, however, that the goal of this paper is methodological – that is, it aims to discuss *how* to answer these questions, not what the answers could be for some particular tradition. The case of ancient Chinese philosophy is used here to illustrate these methodological considerations, and any apparent answers to the three questions in the following should be regarded as illustrations (and even if one would consider them as answers, then they are provisional answers at most).

1. The Concept of TRUTH

The concept of TRUTH (or TRUE, but that is the same concept) is our understanding of the notion that we refer to with the word 'truth,' and *having* the concept of TRUTH is *having* a mental understanding of that notion. By implication, whether someone (or some group) has or had this concept is a psychological question, but unlike many other psychological questions, it can be answered without empirical research on the people involved if there is good reason to belief that the concept of TRUTH is universal. According to Anna Wierzbicka and Cliff Goddard it is,[7] but there are other reasons to believe that TRUTH is universal as well.

Anna Wierzbicka and associates have been attempting to identify *semantic primes* in a research program called *Natural Semantics Metalanguage* (NSM) that

[7] Cliff Goddard, "The Search for the Shared Semantic Sore of All Languages," in *Meaning and Universal Grammar: Theory and Empirical Findings*, ed. Cliff Goddard and Anna Wierzbicka, Volume 1 (John Benjamins, 2002), 5-40.

spans over three decades.[8] Semantic primes are both primitive and universal (and universally primitive), meaning that they cannot be analyzed or paraphrased in any simpler terms, and that they have lexical equivalents (either one or multiple) in all languages (but such lexical equivalents can be polysemous, and there are other complications; see next section). TRUE is one of the semantic primes identified.[9] This means, that according to NSM, not just the concept TRUE or TRUTH is universal (and universally primitive), but also that ·truth· is universal.

There are reasons, however, not to take NSM for granted. For any candidate prime, showing that it cannot be analyzed or paraphrased in any simpler terms in any language – the first criterion of prime-ness – and that it has lexical equivalents in any language – the second criterion – would require a book length study at least, but typically, in the NSM literature, primes are posited and defended within the space of pages.[10] These positings and defenses *seem* to be based on extensive knowledge of language, but remain extremely opaque, and often evoke the suspicion of armchair speculation (or even of being driven by the theory they are supposed to support more than by available data). Furthermore, even if more extensive research would show that TRUE/TRUTH is universal *now*, that does not imply that it always has been. The concept may have become universal fairly recently under the influence of Western cultural dominance, for example.

What we need to know to answer the question whether the ancient Chinese, for example, had the concept of TRUTH is not just whether that concept is *contingently* universal now, but whether it is *necessarily* universal. To answer *that* question, we need to better understand what it means to have the concept of TRUTH first. According to Donald Davidson, TRUTH

> is as fundamental a concept as any we have, for without it we would have no concepts at all. The reasoning is simple: to have a concept is to judge that certain things fall under it, and others don't. To judge that something is, say, lavender, is to hold it to be true that that thing is lavender. To have any propositional attitude requires knowing what it would be for the proposition entertained to be true. Our conviction that there is a way things are however we may think they

[8] See for example: Anna Wierzbicka, *Semantic Primitives* (Frankfurt: Athenäum, 1972); Anna Wierzbicka, *Semantics: Primes and Universals* (Oxford: Oxford University Press, 1996); Goddard, "Search;" Cliff Goddard, "The Natural Semantic Metalanguage Approach," in *The Oxford Handbook of Linguistic Analysis*, ed. Bernd Heine and Heiko Narrog (Oxford: Oxford University Press, 2010), 459-484.

[9] Goddard, "Search."

[10] See, for example, Wierzbicka, *Semantics.*

> are depends on our having the concept of truth, and this is the same as having the concept of an objective reality.[11]

And by implication, "without a grasp of the concept of truth, not only language, but thought itself, is impossible."[12]

Much of Davidson's writing about the concept of TRUTH is related to his controversial theory that having beliefs requires having the concept of BELIEF, which in turn requires having the concept of TRUTH.[13] What is (relatively, at least) uncontroversial, however, is Davidson's insight that having the concept of TRUTH is understanding that there is a difference between what is the case and what is not, and that having the concept of TRUTH is inseparable from having a rather large number of related concepts including both neighbors such as OBJECTIVITY and JUSTIFICATION and contra(dicto)ry concepts such as FALSEHOOD, ERROR, and MISTAKE.[14]

According to John Flavell, children learn to distinguish appearance from reality between the ages of 3 and 4 or 5.[15] The research he reports on depends on linguistic interaction with children, however, which may set the bar too high. More language-independent, observational research has shown that virtually all children start pretend play before the age of 2, and that they are perfectly capable of separating pretense from truth.[16] The ability to distinguish pretense from truth is the ability to distinguish what is (really) the case from what is not, and that ability requires the concept of TRUTH. Therefore, virtually all children develop the concept of TRUTH before the age of 2. Of course, that doesn't imply that 2-year-olds have a *word* for TRUTH (or something similar); that would be confusing TRUTH and ·truth·. Again, one can have a concept without having a word for it. A concept is psychological; it is an ability to make (and understand) a distinction, and the relevant distinction in case of TRUTH is learned at a very early age, well

---

[11] Donald Davidson, "Intellectual Autobiography," in *The Philosophy of Donald Davidson*, ed. Lewis Edwin Hahn (Chicago: Open Court, 1999), 65-66.

[12] Donald Davidson, "Truth Rehabilitated," in *Truth, Language, and History* (Oxford: Oxford University Press, 2005), 16.

[13] Donald Davidson, "Rational Animals," in *Subjective, Intersubjective, Objective* (Oxford: Oxford University Press, 2001), 95-105.

[14] Whether JUSTIFICATION really is a neighbor of TRUTH is debatable, but even if it is not, it is sufficiently close to cause confusion of the two concepts. See section 3.

[15] John Flavel, "The Development of Children's Understanding of False Belief and the Appearance-Reality Distinction," *International Journal of Psychology* 28.5 (1993): 595-604.

[16] Angeline Lillard, Ashley Pinkham, and Eric Smith, "Pretend Play and Cognitive Development," in *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, ed. Usha Goswami (Chichester: Wiley-Blackwell, 2011), 285-311.

before children develop the linguistic tools necessary to reflect on that distinction or even to name it.

The claim that some people or tradition of thought did or does not have the concept of TRUTH is the claim that they did not understand the notions of falsehood, error and mistake; it implies that they did not understand the difference between appearance and reality, or between what is the case and what is not. Aside from the utter implausibility of this claim for *any* people/tradition, there is abundant textual evidence in the case of ancient Chinese philosophy showing that they understood the difference between what is the case and what is not, and thus had the concept of TRUTH. Consider, for example, the following fragment from the *Han Fei Zi* 韓非子:

> 言之為物也以多信，不然之物，十人云疑，百人然乎，千人不可解也。
> Sayings/words are things that are believed because many endorse them. Concerning something that is not *ran* (true?), if ten people say it there is still doubt, if a hundred people say it is considered *ran*, if a thousand people say it it cannot be rejected.[17]

Regardless of whether 然 *ran* is to be translated (here) as 'true' or as one of its neighbors such as 'objective,' 'justified,' or 'the case,' this sentence could not have been written – or even thought – by someone who did not have the concept of TRUTH.

2. Words for ·Truth·

As mentioned in the previous section, according to Natural Semantics Metalanguage (NSM), ·truth· or ·true· is universal: all languages have one or more words, morphemes, or expressions that express TRUTH. Some of these may be polysemous, however, meaning that they only express TRUTH in certain contexts, and recognizing and identifying ·truth· is further complicated by language change, by terminological differences between schools and philosophers, and by opaque compounds and expressions. (The latter kind of complication also occurs in English. For example, 'true' can also mean something like 'genuine,' in which case 'true' is not ·true·.)

W.V.O. Quine called the common idea that there are one-to-one semantic relations between words in different languages the 'myth of the museum.'[18] The

[17] §48:11. All references to Chinese texts (and the paragraph numbers in those references) in this paper are references to the Chinese Text Project edition(s) available at http://ctext.org/. All translations are my own.

[18] W.V.O. Quine, "Ontological Relativity," in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 26-68. Bryan van Norden used the term 'lexical fallacy'

myth may *seem* to be true in the case of ·truth· in Indo-European languages, but it is a myth nevertheless: a language can include ·truth· without having a *single* word for TRUTH, and without having a word that means TRUTH in *all* contexts. At least hypothetically, it is even possible that a language has no word for ·truth· at all (if NSM is wrong). Keeping these complications in mind, how do we recognize and identify ·truth·?

The most obvious identification criterion for ·truth· is: "A word or expression in some language is ·truth· iff it expresses the concept of TRUTH". However, as mentioned above, one cannot have the concept of TRUTH without having a number of related and adjacent concepts including, for example, OBJECTIVITY, and these interconnected concepts cannot be easily separated from each other. If we know that "*ruuv teeh*" means "fire is hot" in some alien language, and that '*tche*' means *something like* 'true,' then that doesn't settle the correct translation of "*ruuv teeh tche.*" That sentence could mean "it is true that fire is hot" or "it is objective (-ly the case) that fire is hot" (among other options), and these alternative translations are not equivalent (the second expresses independence from perspective or point of view, for example). And lacking evidence for which TRUTH-like concept exactly '*tche*' expresses, one is not justified to identify it as ·truth·. Hence, we need some additional criterion or criteria to distinguish ·truth· from its neighbors.

Firstly, ·truth· is attributed to sentences, propositions, beliefs or something very similar, taking relevant grammatical differences between languages into account. The sentences (*etc.*) that are judged to be true or not can be fairly simple as in the case of predicate-subject sentences, or very complex as in the long, compound propositions that are needed to represent theories. In case of some languages such as classical Chinese predicate-subject sentences can be expressed by means of a single word or character, which may be a source of confusion. For example, if context specifies the subject *x* of predication, then the one-character sentence '白' *bai* has the propositional content 'white (*x*).' Because of this feature of classical Chinese, a two-character sentence 「厶白」, in which the character 厶 (a semantic variant of 某, which means 'some') is a placeholder for a candidate character for ·true·, is ambiguous if 厶 can also be interpreted as an adverb. The sentence 「厶白」 would then be interpretable either as "It is true that *x* is white" or as "*x* is truly white." In the latter case (*i.e.* in its adverbial use)厶 is not

to refer to the related mistake of assuming that some tradition doesn't have a concept of *x* or views about *x* because it doesn't have a single word for *x*. See: Bryan Van Norden, *Virtue Ethics and Consequentialism in Early Chinese Philosophy* (Cambridge: Cambridge University Press, 2007).

attributed to a sentence, and therefore, is not ·true· (but is more similar to the use of 'truly' in English to mean 'genuinely' or 'really'), but that doesn't mean that ⏃ is not ·true· either in its other use (*i.e.* as a sentence-level operator or a property of a proposition).

Secondly, ·truth· is *disquotational,* but its neighbors are not. According to (a variant of) Tarski's famous T-schema:

TS it is true that $p \leftrightarrow p$

TS holds for ·true·, but not for its conceptual neighbors. If it is the case that *p*, then it is the case that *p* is true, and the other way around – that is what TS means. However, *p* may be objective or justified without it being the case that *p*, and/or the other way around. Consequently, contrary to ·true·, these conceptual neighbors are not disquotational. (Note that in the above example, ⏃ in its adverbial use is not disquotational, which is another reason why it is not ·true·.)

It must be emphasized that the notion of disquotation should not be confused with theories of truth that claim that disquotation defines truth or that disquotation is all there is to say about truth (*i.e.* deflationism or minimalism about truth; see next section). There may or may not be much more to say about the property of *truth*, but that is not the issue here. Rather, I'm merely making the much more uncontroversial claim that ·truth· (or ·true·) can be recognized by its satisfaction of TS (taking the symbol '↔' to represent nothing but material equivalence). That this is an uncontroversial claim follows from the fact that nearly all contemporary theories of truth accept some form of TS.

With these two additions, the following identification criterion for ·truth·/·true· can be formulated:

> A word or expression in some language is ·truth·/·true· in some (kind of) use and context iff,
>
> in that (kind of) use and context
>
> a) it is most plausibly interpreted as expressing the concept of TRUTH,
>
> b) it is attributed to a sentence, proposition, belief, or something very similar,
>
> and
>
> c) it is disquotational.

This criterion is admittedly imprecise. It takes the ability to recognize whether a word expresses TRUTH for granted, for example. And it leaves open many questions with regards to 'use and context.' How often (between once and always) should a word satisfy this criterion to be considered ·truth·? What specifies the *kind* of use or context? And so forth. None of this is problematic,

however. It would be if the criterion would misidentify some words as ·true·, but I have been unable to find such misidentifications for English. Most likely candidates would be the English expressions "it is the case that" and "it is a fact that," which satisfy (b) and (c), and possibly also (a). If they do also satisfy (a) – and I believe they do, but will not defend that belief here – then indeed they are examples of ·true·, but I doubt that this identification as such would (or should, at least) be controversial. (On the other hand, "there is a fact that" does not satisfy (c) as it involves an ontological commitment to facts that is absent in "it is true that.")

In case of ancient Chinese philosophy, several candidates for ·truth· have been suggested in the literature. Most prominent are 真 *zhen*, 實 *shi,* 是 *shi*, 然 *ran*, 當 *dang,* and perhaps 可 *ke.* The last is used in various texts as an apparent property of sentences. For example, the opening sentences of Gong Sun Long's 公孫龍 *Bai Ma Lun* 白馬論 are:

> 「白馬非馬」，可乎？曰：可。 Is "a white horse is not a horse" admissible (*ke*)? It is (*ke*).

Here 可*ke*, which (following Hansen)[19] I translated as 'admissible' appears to be a property of the sentence 「白馬非馬」. It *can* be interpreted as meaning TRUE in this context, and in that case, it would probably be disquotational, but it is doubtful whether that interpretation is correct. It depends on whether the reply 「可」 should be taken to imply an affirmation of 「白馬非馬」rather than just of its admissibility and it is by no means certain that it is intended as such.

Most of the other characters mentioned are more likely candidates for truth. Chris Fraser makes a case for 當 *dang* in the context of Mohism;[20] Alexus McLeod argues for 實 *shi,* 是 *shi*, and 然 *ran* in the writings of Wang Chong 王充;[21] and Wai Chun Leong argues for 然 *ran*.[22] Textual ambiguities make it very difficult to judge whether these indeed satisfy the above identification criterion. They can all be defended as expressions of TRUTH, but as argued above, TRUTH is not easily separated from its neighbors such as OBJECTIVITY, and all of them can be just as easily interpreted as expressing some adjacent concept.

---

[19] Hansen, "Chinese Language."

[20] Chris Fraser, "Truth in Moist Dialectics," *Journal of Chinese Philosophy* 39.3 (2012): 351-368.

[21] Alexus McLeod, "Pluralism about Truth in Early Chinese Philosophy: a Reflection on Wang Chong's Approach," *Comparative Philosophy* 2.1 (2011): 38-60. For an opposing point of view, see: Lajos Brons, "Wang Chong, Truth, and Quasi-pluralism," *Comparative Philosophy* 6.1 (2015): 129-148.

[22] Wai Chun Leong, "The Semantic Concept in Truth in Pre-Han Chinese Philosophy," *Dao* 14.1 (2015): 55-74.

In practice, only the third sub-criterion – that of disquotationality – can tell us whether a candidate word/character is ·truth·, but as Leong points out, there is very little (if any) textual evidence for disquotationality. A word/character 厶 is disquotational if asserting that '*p*' is 厶 is asserting (that) *p* and *vice versa* (ignoring the fact that there may be a pragmatic difference between '*p*' and "'*p*' is 厶"), but nothing resembling this pattern occurs. The closest Leong could find is Mencius' 孟子 reply 「然」to a factual question in Gao Zi II 告子下 §22, but a better example of this kind of use is the following fragment from the chapter Gong Sun Chou II 公孫丑下 §18:

> 曰：「使管叔監殷，管叔以殷畔也，有諸？」曰：「然。」 Is it the case that [the duke of Zhou] sent Guan Shu to supervise [the state of] Yin but that Guan Shu with Yin rebelled? It is (*ran*).

Here 然 *ran* is used in reference to the factual description (*i.e.* a proposition) in the question. Mencius' reply seems to be short for "it is *ran* that the duke of Zhou sent Guan Shu to supervise Yin but that Guan Shu with Yin rebelled," and if that analysis is correct, this would probably be the best example of *ran* as a disquotational property of propositions or sentences, and therefore, as ·truth·/·true·. However, this is not the only possible analysis. Similar occurrences of *ran* can be found throughout *the Analects* and the *Meng Zi*,[23] and in most cases it can be translated as an affirmation comparable to "Yes" in English. If *ran* (in this kind of use) merely affirms, it is not a property of a proposition, and thus not ·truth·/·true·. The interpretation of *ran* as "Yes" rather than ·true· runs into trouble, however, in case of Mencius' reply (to another question) 「否，不然」("No, not *ran*") in *Wan Zhang* 萬章 I:§9. It is hard to make sense of this reply otherwise than as "No, not true" (or as "No, not the case," but as argued above, the English expression "is the case" is a form of ·true·).

That 然 *ran* almost certainly is ·truth·/·true· in some (con)texts does not imply that the other words/characters mentioned above are not (in other (con)texts), but as the case of ancient Chinese philosophy is mere illustration here, it suffices to show that there is at least one word/character for ·truth·/·true· in classical Chinese.

3. The Property of *Truth*, and Theories of *Truth*

The property of *truth* is the feature or collection of features (if that or those exist) that all true sentences/propositions/beliefs share and that all false ones lack. Most

[23] See, for example, in *the Analects*: Wei Ling Gong 衛靈公 §42, Yang Huo 陽貨 §7, Wei Zi 微子 §6; in *Meng Zi*: Teng Wen Gong I 滕文公上 §2 and 4; Gao Zi I 告子上 §3.

theories of truth try to define TRUTH in such a way that the definition captures the property of *truth*. This property is what *makes* some sentence or proposition true, or *by virtue of which* a sentence or proposition is true. Deflationists and primitivists about truth deny that there is such a property, while correspondentists, coherentists, and other substantivists claim there is. Correspondentism and deflationism (or 'minimalism') are the two dominant positions in the contemporary debate.

According to most correspondentists, '*p*' is true if and only if it corresponds with a fact that *p*. This view needs to be distinguished from the more general idea that a sentence or proposition is true if it corresponds with the way things are. 'Corresponding with the way things are' is not a property of *truth*, but is just another way of expressing TRUTH: "the way things are is such that *p*" is synonymous with "it is true that *p*," "it is the case that *p*," and a number of further equivalent expressions. 'Correspondence with a fact that *p*,' on the other hand, means that there is a fact that *p*, and that this fact makes '*p*' true (and thus involves an ontological commitment to facts). Another way to bring out the difference is to focus on the correspondence relation. For correspondentism, this is a relation between discrete truthmakers (such as facts) and truthbearers (sentences or propositions), such that one specific truthmaker makes one specific truthbearer true. The correspondence relation in 'corresponding with the way things are,' 'corresponding with the world,' or Aristotle's "to say of what is that it is" is of an entirely different nature because it does not pick out specific truthmakers, and thus does not specify what makes individual true sentences or propositions true. All it does, is attempt to express what we mean with ·truth·. It is for this reason, that Aristotle's remark "hardly qualifies as a theory of truth" (as I stated in the first paragraph of the introduction).

Substantive theories of truth claim that there is a property of *truth* in virtue of which true sentences/propositions/beliefs are true or that makes them true. A theory about what makes some statement true is deceptively similar, however, to a theory about what justifies one to believe that statement, or about what makes that statement reliable.[24] Therefore, to judge whether some particular theory is a

[24] This problem (like most of the problems discussed in this paper) is not typical of Chinese philosophy: in Western philosophy TRUTH is also often confused with its neighbors. Much of this confusion seems to stem from the common idea that truth is a norm, but as Davidson pointed out, "we do not aim at truth but at honest justification." When we say that we want our statements or beliefs to be true, what we want is overwhelming evidence or an irrefutable argument, but that is justification, not truth. And according to Davidson, we *cannot* ask for more than that; "it makes no sense to ask for more." (The two quotes are from: Donald

theory about the property of *truth*, we need to establish with sufficient certainty that that theory is about what makes some statement(s) true rather than justified, reliable, objective, and so forth.

In case of ancient Chinese philosophy, a few theories have been suggested as possible theories of truth, but none of the suggestions that I am aware of is convincing. For example, Chris Fraser suggests Mo zi's 墨子 three standards (*Fei Ming Shang* 非命上 §2) as a theory of truth,[25] but this is not the most plausible interpretation. The three standards – *basis* (本: based on the deed of the ancient sage-kings), *source* (原: hearing and sight of common people), and *use* (用: beneficial to the state and the people) – may make some statement justified, assertable, reliable, plausible, acceptable, believable, appropriate, and so forth, but it is hard to believe that Mo zi claimed that they *make* a statement true. The context of Mo zi's argument and similar arguments in, for example, *Han Fei Zi* 韓非子§30*ff* is pragmatic. What matters is that the actions and policies of the state are based on reliable information. Moreover, even in the strongest interpretation, Mo zi's three standards or Han Fei's seven techniques 七術, are more plausibly interpreted as being about justification than about truth. In Mo zi's view, the three standards may very well justify believing that some statement is true, or accepting it as true, but 'justification to believe that true' or 'justification to accept as true' is justification, not truth, and certainly not synonymous with 'true.'

To attribute a theory of *truth* to the ancient Chinese (or to some ancient Chinese philosopher), we would need to find an account of what makes some statement true that cannot be interpreted more plausibly as being about what makes it justified, acceptable, reliable, or some other neighbor of TRUTH. I'm not aware of any such account, and I doubt that there is one. This then, would mean that the ancient Chinese didn't have a theory (or theories) of *truth* (but they did have theories of justification, and thus epistemology). The same may very well be the case for many other non-Western traditions of thought.

This conclusion should not be reason for surprise, however. It has taken Western philosophers some time to reflect on their obsession with TRUTH and *truth*, but nowadays deflationism (which denies that there is a substantive property of *truth*) is one of two mainstream accounts of truth (correspondentism is the other). *If* deflationists (and primitivists, which agree with deflationists on this point) are right, then the Chinese lack of a theory (or theories) of truth is less

---

Davidson, "Reply to Pascal Engel," in *The Philosophy of Donald Davidson*, ed. Lewis Edwin Hahn (Chicago: Open Court, 1999), 461.)

[25] Fraser, "Truth."

strange than the existence of such theories in Western philosophy; then Chinese philosophy avoided a dead end from which Western philosophy is now finally, reluctantly returning. But even if they are wrong, the fact that deflationism and primitivism are defensible accounts of truth shows that it is not a defect of some philosophical tradition not to attempt to define TRUTH or to theorize about *truth*.

What must be emphasized, however, is that the apparent lack of theories about *truth* or definitions of TRUTH in Chinese philosophy does not imply an (even implicit) adherence to a variety of deflationism or primitivism. What defines the latter is that they claim that TRUTH cannot be defined and/or that there is no substantive property of *truth*, and either claim can only be made in the context of explicit theorizing about *truth*. In other words, even though deflationism and primitivism deny that there is a property of *truth*, that denial itself is a theory of *truth*, and consequently, not having a theory of *truth* does not suffice for the classification as deflationist or primitivist.

Conclusion: Recognizing 'Truth'

The debate about truth in ancient Chinese philosophy is based on the assumption that it concerns a single question, thus confusing terms, concepts, and theories of truth. It is essential, however, to distinguish the property of *truth* (the shared feature of true statements that makes them true), the concept of TRUTH (our mental understanding of truth), and ·truth· (words used to express TRUTH; *i.e.* equivalents of the English word 'truth'). When this distinction is made, the question about truth in Chinese philosophy no longer is one question, but three questions. These three questions are the same for any non-Western tradition of (philosophical) thought:

1) Does that tradition have a theory about the property of *truth*?

2) Does that tradition have the concept of TRUTH?

3) Does that tradition's language include ·truth·?

This paper did not intend to give (definitive) answers to these questions for the case of ancient Chinese philosophy, but to reflect on *how* such questions should be answered for any tradition. Hence, the concern of this paper is methodological rather than topical.

Because in all likelihood all (normal) human beings over the age of 2 have the concept of TRUTH, the answer to question (2) is always "Yes," regardless of the tradition investigated. Having a concept does not necessarily imply having a single word to express it, however, and identifying words for ·truth· or ·true· is complicated by neighboring concepts: if a word in some language can equally

plausibly interpreted as 'objective' or 'justified' then the interpreter is not justified to just assume it means 'true.' ·Truth· can be distinguished from its neighbors, however, because it has a characteristic its neighbors lack: disquotationality. If it is the case that *p* is true, then it is the case that *p*, and the other way around. Armed with this criterion, ·truth·/·true· can be distinguished from its neighbors. In the case of ancient Chinese 然 *ran* is ·truth·/·true· in at least certain uses and possibly there are other words/characters for ·truth· as well.

A theory of truth is a theory about the property of *truth*. It either specifies the nature of that property – thus giving criteria something must satisfy to be true – or it denies that there is such a property. It is essential to distinguish theories of *truth* from theories of *justification* (and theories of *objectivity*, and so forth). TRUTH and JUSTIFICATION are adjacent concepts, but they are not identical, and a statement may be justified without being true and/or the other way around. Consequently, to determine whether some particular theory is a theory of *truth* it needs to be made sufficiently plausible that it isn't better understood as a theory of *justification* (or other conceptual neighbor). There appear to be no such theories in ancient Chinese philosophy, and it may very well be the case that the same is true for many (if not most) other non-Western traditions. If deflationists and primitivists are right that there is no property of *truth* then this is hardly surprising, however. Then Western theorizing about *truth* is more peculiar than the absence thereof elsewhere.

SEMANTIC EPISTEMOLOGY REDUX: PROOF AND VALIDITY IN QUANTUM MECHANICS

Arnold CUSMARIU

ABSTRACT: Definitions I presented in a previous article as part of a semantic approach in epistemology assumed that the concept of derivability from standard logic held across all mathematical and scientific disciplines. The present article argues that this assumption is not true for quantum mechanics (QM) by showing that concepts of validity applicable to proofs in mathematics and in classical mechanics are inapplicable to proofs in QM. Because semantic epistemology must include this important theory, revision is necessary. The one I propose also extends semantic epistemology beyond the 'hard' sciences. The article ends by presenting and then refuting some responses QM theorists might make to my arguments.

KEYWORDS: Richard Feynman, J.M. Jauch, paradigm shift, proof, quantum mechanics, semantic epistemology, validity

1. Introduction

In an earlier article,[1] I presented and defended definitions of semantic evidence in science and in mathematics as part of a general semantic approach in epistemology. The first clauses of these definitions read as follows ('SL' and 'ML' abbreviate 'scientific language' and 'mathematical language,' respectively):

> (SES1) Where *z* is a wff of a scientific language *SL*, *z*-is-evident-in-*SL* for S =Df (i) There is a derivation-in-*SL* of *z* from true-in-*SL* instrumental-accuracy-law-sentences-of-*SL* and initial-condition-sentences-of-*SL*.
>
> (SEM1) Where *z* is a wff of a mathematical language *ML*, *z* is evident-in-*ML* for a person S =Df (i) There is a derivation-in-*ML* of *z*.

Three assumptions were made that seemed obvious at the time:

Assumption 1: SES(i) and SEM(i) are entitled to employ the same concept of derivability from standard logic. This assumption was made so that a semantic evidence predicate could be formulated along deductivist lines for both science and mathematics.[2]

[1] Arnold Cusmariu, "Toward a Semantic Approach in Epistemology," *Logos & Episteme. An International Journal of Epistemology* III, 4 (2012): 531-543.

[2] Cusmariu, "Toward a Semantic Approach," 533.

Assumption 2: The same concept of derivability from standard logic holds across all mathematical disciplines and their respective languages. This assumption was proved by Bertrand Russell and A.N. Whitehead in *Principia Mathematica*.
Assumption 3: The same concept of derivability from standard logic holds across all scientific disciplines and their respective languages. Physics has taken this assumption for granted ever since Newton's derivation of Kepler's Laws from his own.

I will show that Assumption 3 is not true for quantum mechanics (QM), likewise Assumption 1. Using as case study a proof by J.M. Jauch in his *Foundations of Quantum Mechanics*,[3] I show that standard concepts of validity applicable to proofs in mathematics and in classical mechanics are inapplicable to proofs in QM; therefore, SES1 must be revised to include this important theory. The one I propose also extends semantic epistemology beyond the 'hard' sciences. The article ends by showing the inadequacy of some responses QM theorists might make to my arguments and suggests that the unavailability of standard logic is a reason QM may represent a paradigm shift. Assessing what exactly that entails is beyond the scope of this article.

2. Case Study Preliminaries

Mathematical proofs work by establishing logical links to previous results. A *reductio ad absurdum* proof, which Pythagoras used to show that √2 is irrational, goes about it in a special way. Here a proposition A is proved by showing that its negation, $\sim A$, leads to contradiction, C, from which A follows because contradictions are false. G.H. Hardy thought the *reductio* was "one of a mathematician's finest weapons."[4]

Though QM is an empirical theory – as is classical mechanics – there have also been efforts to prove results by purely logical means, including the *reductio* method. In his book, J.M. Jauch proved the following by *reductio*:

> Proposition 1: Every dispersion-free state is pure.

Jauch proves Proposition 1 by deriving a contradiction from its negation,

> Proposition 2: There is a dispersion-free and mixed state.

Here is the language of his proof:[5]

[3] Josef M. Jauch, *Foundations of Quantum Mechanics* (Reading, Massachusetts: Addison-Wesley, 1968). See also Constantin Piron, *Foundations of Quantum Mechanics* (London: Benjamin, 1976).
[4] G.H. Hardy, *A Mathematician's Apology* (Cambridge: Cambridge University Press, 1940), 94.
[5] Jauch, *Foundations of Quantum Mechanics*, 115.

> Suppose the state p is a mixture. Then there exist two different states p_1 and p_2, as well as two positive numbers λ_1 and λ_2 such that $\lambda_1 + \lambda_2 = 1$ and $p = \lambda_1 p_1 + \lambda_2 p_2$. Since the two states p_1 and p_2 are different from one another, there exists a proposition a such that $p_1(a) \neq p_2(a)$ (cf. Property 5b of Section 6-3). Since the states are dispersion-free, there are two possibilities only: $p_1(a) = 1$, $p_2(a) = 0$ or $p_1(a) = 0$, $p_2(a) = 1$. In either case we have $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$. Thus the state is not dispersion-free, contrary to the assumption. This proves the proposition.

In a tradition going back to Euclid, Jauch presents only information he thinks is sufficient to make it apparent that the argument is logically correct (valid) – or, as Fermat famously put it, to 'compel belief' (*forcer á croire*).[6] Proving that the argument is valid, however, is another matter entirely.[7] A standard way of doing that in logic is by means of a formal proof of validity (FPV), which entails making argument steps explicit all the way to the conclusion and stating the rules of logic used to derive inferred steps.

3. A Formal Proof of Validity of Jauch's *Reductio*

Jauch's Proposition 1 is a universally quantified material conditional of first-order logic and may be symbolized as

(P1) $(x)(Fx \rightarrow Gx)$,

while its negation, Proposition 2, is an existentially quantified conjunction of first-order logic and may be symbolized as

(P2) $(\exists x)(Fx \,\&\, \sim Gx)$.

This construal of P1 and P2 is reasonable because Jauch uses the terms 'every' and 'there exist,' which denote universal and existential quantifiers, respectively, and has them apply to states characterized as dispersion-free, pure or mixed. This suggests that states are the objects of quantification in P1 and P2

[6] Paul Tannery and Charles Henry, eds., *Œvre de Fermat* (Paris: Blanchard, 1891-1912), Vol. II, 483.

[7] The idea that syntactic validity is proved by reference to rules of inference is due to Aristotle. Unfortunately, his list of valid syllogisms, which effectively function as such rules, turned out to be inadequate for general mathematical purposes – including the FPV of Jauch's *reductio* below – and there the matter rested until Frege's discovery of quantification. For more on these issues, see Arnold Cusmariu, "A Methodology for Teaching Logic-Based Skills to Mathematics Students," *Symposion: Theoretical and Applied Inquiries in Philosophy and Social Sciences* 3, 3 (2016): 259-292, esp. 259-261.

rather than the properties of being dispersion-free, pure and mixed. Quantification in P1 and P2 is first, not second order.[8]

However, strictly speaking, the negation of P1 is not P2 but rather

(P3) $\sim(x)(Fx \rightarrow Gx)$.

An FPV from P1 to P2 is needed before the full argument can get under way. Working with logical-form versions of Propositions 1 and 2 is sufficient for this purpose.

Steps	Justification
(1) $\sim(x)\Phi x \equiv (\exists x)\sim\Phi x$	*Quantifier Negation Law* (QN)
(2) $\therefore \sim(x)\Phi x \rightarrow (\exists x)\sim\Phi x$	From 1 by *Material Equivalence*, *Simplification* (Simp.)
(3) $\therefore \sim(x)(Fx \rightarrow Gx) \rightarrow (\exists x)\sim(Fx \rightarrow Gx)$	From 2 by *Substitution*
(4) $\therefore (\exists x)\sim(Fx \rightarrow Gx)$	From P3, 3 by *Modus Ponens* (MP)
(5) $\therefore (\exists x)\sim(Fx \rightarrow Gx) \rightarrow (\exists x)\sim(\sim Fx \vee Gx)$	From 4 by *Material Implication*
(6) $\therefore (\exists x)\sim(\sim Fx \vee Gx)$	From 4, 5 by MP
(7) $\therefore (\exists x)\sim(\sim Fx \vee Gx) \rightarrow (\exists x)(\sim\sim Fx \,\&\, \sim Gx)$	From 6 by *De Morgan's Theorem*
(8) $\therefore (\exists x)(\sim\sim Fx \,\&\, \sim Gx)$	From 6, 7 by MP
(9) $\therefore (\exists x)(\sim\sim Fx \,\&\, \sim Gx) \rightarrow (\exists x)(Fx \,\&\, \sim Gx)$	From 8 by *Double Negation*
(P2) $\therefore (\exists x)(Fx \,\&\, \sim Gx)$	From 8, 9 by MP

The FPV of Jauch's argument can now proceed.

Steps	Justification
$\therefore$ (1) p is a dispersion-free and mixed state.	From Proposition 2 by *Existential Instantiation*
$\therefore$ (2) p is a dispersion-free state.	From 1 by Simp.
(3) If p is a dispersion-free and mixed state, then p consists of dispersion-free states $p_1 \neq p_2$ such that $p = \lambda_1 p_1 + \lambda_2 p_2$ for positive numbers $\lambda_1 + \lambda_2 = 1$.	Assumption
$\therefore$ (4) p consists of dispersion-free states $p_1 \neq p_2$ such that $p = \lambda_1 p_1 + \lambda_2 p_2$ for positive numbers $\lambda_1 + \lambda_2 = 1$.	From 1, 3 by MP
$\therefore$ (5) If p consists of dispersion-free states $p_1 \neq p_2$ such	From Property 5b of Section 6-3,

[8] Jauch's identity conditions for states suggest a first-order interpretation: "Two states are identical if the relevant conditions in the preparation of the state are the same." (Jauch, *Foundations of Quantum Mechanics*, 92.) The ontological status of states does not affect the arguments presented below.

that $p = \lambda_1\ p_1 + \lambda_2\ p_2$ for positive numbers $\lambda_1 + \lambda_2 = 1$, then, $p_1(a) \neq p_2(a)$ for some proposition *a*.	by *Universal Instantiation*[9]
∴ (6) $p_1(a) \neq p_2(a)$ for some proposition *a*.	From 4, 5 by MP
(7) If $p_1(a) \neq p_2(a)$ for some proposition *a*, then, either $p_1(a) = 1$ and $p_2(a) = 0$; or, $p_1(a) = 0$ and $p_2(a) = 1$.	Assumption
∴ (8) Either $p_1(a) = 1$ and $p_2(a) = 0$; or $p_1(a) = 0$ and $p_2(a) = 1$.	From 6, 7 by MP
(9) If Either $p_1(a) = 1$ and $p_2(a) = 0$; or $p_1(a) = 0$ and $p_2(a) = 1$, then, $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$.	Assumption
∴ (10) $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$.	From 8, 9 by MP
∴ (11) If $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$, then *p* is not a dispersion-free state.	From the definition of 'dispersion-free' state.[10]
∴ (12) *p* is not a dispersion-free state.	From 10, 11 by MP
∴ (13) *p* is a dispersion-free state and *p* is not a dispersion-free state.	From 2, 12 by *Conjunction*
∴ (14) ~(*p* is a dispersion-free state and *p* is not a dispersion-free state.)	From (13) by the *Law of Non-Contradiction*
∴ (15) If ~(*p* is a dispersion-free state and *p* is not a dispersion-free state), then, it is not the case that *p* is a dispersion-free and mixed state.	From (1) and (14), shortcut[11]
∴ (16) It is not the case that *p* is a dispersion-free and mixed state.	From (14), (15) by MP
∴ (17) It is not the case that there is a dispersion-free and mixed state.	From (16) by *Existential Generalization*
∴ (18) Every dispersion-free state is pure. Q.E.D.	From (17) by QN[12]

[9] Jauch, *Foundations of Quantum Mechanics*, 94. This is one of several properties postulated.
[10] Jauch, *Foundations of Quantum Mechanics*, 114.
[11] We omit the laborious process of deriving (15) to avoid cluttering the text. The points made below do not require a full expansion of the argument for step (15).
[12] An FPV of (18), which has the form $(x)(Fx \rightarrow Gx)$, from (17), which has the form $\sim(\exists x)(Fx \,\&\, Gx)$, is analogous to the previous FPV and may be omitted.

4. Syntactic and Semantic Validity

The above is a proof of validity in the syntactic sense, according to which an argument is syntactically valid if and only if all inferred steps are derived according to rules of logic. A concept distinct from syntactic validity is semantic validity, according to which an argument is semantically valid if and only if its conclusion is true if the premises are true for any truth-functional interpretation of premises and conclusion. That is, an argument from premises $\{p_1, p_2, p_3 \ldots p_n\}$ to conclusion c is semantically valid just in case the corresponding material conditional $(p_1 \& p_2 \& p_3 \& \ldots p_n) \rightarrow c$ is a tautology.

Let us provide a proof of semantic validity in the context of Jauch's argument. It will be sufficient to do so only for a portion of the argument because it is elementary how the proof can be generalized for the entire argument.

Thus, consider the inference to step (12) from premises (10) and (11):

(10) $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$.

(11) If $\sigma(a) = \lambda_1 + \lambda_2 \neq 0$, then p is not a dispersion-free state.

$\therefore$ (12) p is not a dispersion-free state.

For ease of reference, let us first abbreviate (10) as r; (11) as $r \rightarrow \sim s$; and (12) as $\sim s$. The material conditional corresponding to the argument from (10) and (11) to (12) is $(r \& (r \rightarrow \sim s) \rightarrow \sim s)$. Next we enter this sentence and its components into a truth table configured according to standard semantics for logical connectives.

	1	2	3	4	5	6
1	r	s	$\sim s$	$r \rightarrow \sim s$	$r \& (r \rightarrow \sim s)$	$(r \& (r \rightarrow \sim s)) \rightarrow \sim s$
2	T	T	F	F	F	T
3	T	F	T	T	T	T
4	F	T	F	T	F	T
5	F	F	T	T	F	T

If the argument from (10) and (11) to (12) is semantically valid, then we should find that the material conditional corresponding to this argument, $(r \& (r \rightarrow \sim s) \rightarrow \sim s)$, is a tautology, meaning that column 6 should show only the truth value True, which it does. This completes the proof of semantic validity for the argument from (10) and (11) to (12) and, by implication, Jauch's entire argument.

It is not a coincidence that a syntactically valid argument has turned out to be semantically valid as well. Though there is no need here to address the general

problem of equivalence between syntactic and semantic validity, the following points are relevant to the arguments that will emerge shortly.

First, we note that it is standard to define logical connectives by means of binary truth values as shown in the above truth table. Thus, column 3 shows the definition of negation; column 4 of material implication; and column 5 of conjunction. Though disjunction is not shown, its definition is set by binary truth values in similar fashion. Logical connectives are in general defined by standard truth-table semantics.

Second, standard semantics for logical connectives also define the concept of a tautology.

Third, the fact that standard semantics for logical connectives define the concept of a tautology means that the definition of semantic validity also assumes such semantics.

Fourth, the definition of syntactic validity also assumes standard semantics for logical connectives because: (a) logically compound sentences occur routinely in arguments, certainly mathematical arguments; and (b) rules of logic applied to derive inferred steps assume such semantics.

Point (b) may be obviated by noting that rules of logic such as MP, applied in the FPV above, assume standard semantics for material implication; and by noting that MP works because it is itself a semantically valid argument, meaning that a tautology corresponds to it. A truth table will show that the symbolic sentence corresponding to MP, $(p \,\&\, (p \rightarrow q)) \rightarrow q$, is indeed a tautology. It is apparent that standard semantics for material implication and conjunction must be assumed to show that this sentence is a truth-table tautology.

5. Logical Connectives and QM's Uncertainty Principle

According to truth tables defining standard semantics for logical connectives, the following sentences are tautologies:

(D1) $p \,\&\, (q \vee r) \rightarrow (p \,\&\, q) \vee (p \,\&\, r)$

(D3) $p \rightarrow ((p \,\&\, q) \vee (p \,\&\, \sim q))$

Thus, all eight rows of column 9 linking D1 components in columns 7 and 8 by means of material implication show the truth-value True.

	1	2	3	4	5	6	7	8	9
	p	q	r	$p \,\&\, q$	$p \,\&\, r$	$q \vee r$	1 & 6	4 v 5	7 → 8
1	T	T	T	T	T	T	T	T	T

2	T	T	F	T	F	T	T	T	T
3	T	F	T	F	T	T	T	T	T
4	T	F	F	F	F	F	F	F	T
5	F	T	T	F	F	T	F	F	T
6	F	T	F	F	F	T	F	F	T
7	F	F	T	F	F	T	F	F	T
8	F	F	F	F	F	F	F	F	T

Likewise, all four rows of column 7 linking D3 components in columns 1 and 6 by means of material implication show the truth-value True.

	1	2	3	4	5	6	7
	p	q	$\sim q$	$p \,\&\, q$	$p \,\&\, \sim q$	4 v 5	$1 \rightarrow 6$
1	T	T	F	T	F	T	T
2	T	F	T	F	T	T	T
3	F	T	F	F	F	F	T
4	F	F	T	F	F	F	T

We wish to show that the Uncertainty Principle (UP) of QM is not consistent with the tautological status of D1 and D3. Let us consider them in turn.

UP-D1 Inconsistency: The equivalence

(D) $p \,\&\, (q \vee r) \equiv (p \,\&\, q) \vee (p \,\&\, r)$,

is the rule of replacement *Distribution*, which is a conjunction of material conditionals:

(D1) $p \,\&\, (q \vee r) \rightarrow (p \,\&\, q) \vee (p \,\&\, r)$

(D2) $(p \,\&\, q) \vee (p \,\&\, r) \rightarrow p \,\&\, (q \vee r)$

UP is inconsistent with the tautological status of D1 because it implies that the antecedent of D1, $p \,\&\, (q \vee r)$, can be true but the consequent of D1, $(p \,\&\, q) \vee (p \,\&\, r)$, is false because both disjuncts are false.

Thus, consider the following scenario:

- p is the proposition that the momentum of particle x is in the interval [0, +1/6],
- q is the proposition that the position of particle x is in the interval [–1, +1],
- r is the proposition that the position of particle x is in the interval [+1, +3][13]

Note first that the scenarios described by the three propositions p, q and r, taken singly, are consistent with UP. Second, scenario p & (q v r), the antecedent of D1, is also consistent with UP, and is in fact one of several truth-functional combinations of p, q and r that are consistent with UP.

However, scenario (p & q) v (p & r) is not consistent with UP because it asserts restrictions on simultaneous values of position and momentum of a particle that are not consistent with UP. If UP is true, the two components of the consequent of D1, (p & q) and (p & r), are both false, hence their disjunction is false.[14]

It is important not to misunderstand how this UP-based counterexample to the status of D1 as a tautology arises. It does not arise from the two scenarios taken singly because several truth-table rows for both scenarios show truth values consistent with UP.

Thus, column 7 of the truth table shows three rows where the scenario p & (q v r) has the truth value True:

	1	2	3	6	7
	p	q	r	q v r	1 & 6
1	T	T	T	T	T
2	T	T	F	T	T
3	T	F	T	T	T

Likewise, column 8 of the truth table below shows five rows where the scenario (p & q) v (p & r) has the truth value False:

[13]Scenario details are from a *Wikipedia* article, https://en.wikipedia.org/wiki/Quantum_logic, accessed July 28, 2015.

[14]Note that D2 is consistent with UP, because a conditional is trivially true if it has a false antecedent, which is the case if (p & q) v (p & r) is false according to UP because both disjuncts are false. Note also that 'rejecting D1' is shorthand for "rejecting that D1 is a tautology."

	1	2	3	4	5	8
	p	*q*	*r*	*p* & *q*	*p* & *r*	4 v 5
4	T	F	F	F	F	F
5	F	T	T	F	F	F
6	F	T	F	F	F	F
7	F	F	T	F	F	F
8	F	F	F	F	F	F

Rather, the problem is that there are no rows where both disjuncts of (*p* & *q*) v (*p* & *r*) are false (columns 4 and 5) and *p* & (*q* v *r*) is true (column 7).

	1	2	3	4	5	7
	p	*q*	*r*	*p* & *q*	*p* & *r*	1 & 6
2	T	T	F	T	F	T
3	T	F	T	F	T	T

Nor are there rows where both disjuncts of (*p* & *q*) v (*p* & *r*) are false (columns 4 and 5) and *p* & (*q* v *r*) is true (column 7).

	1	2	3	4	5	6	7
	p	*q*	*r*	*p* & *q*	*p* & *r*	*q* v *r*	1 & 6
4	T	F	F	F	F	F	F
5	F	T	T	F	F	T	F
6	F	T	F	F	F	T	F
7	F	F	T	F	F	T	F
8	F	F	F	F	F	F	F

UP-D3 Inconsistency: UP rejects the tautological status of D3

D3. $p \rightarrow (p \,\&\, q)$ v $(p \,\&\, \sim q)$,[15]

as well because UP blocks the material implication of $(p \,\&\, q)$ v $(p \,\&\, \sim q)$ from p under the following circumstances. Let B_1 and B_2 be Borel sets and p, q and $\sim q$ be the following propositions:

- p is the proposition that a measurement of the momentum of a particle will yield a value in B_1.
- q is the proposition that a (simultaneous) measurement of the position of a particle will yield a value in B_2.
- $\sim q$ is the proposition that a (simultaneous) measurement of the position of a particle will not yield a value in B_2.[16]

Even though p is consistent with UP, nevertheless according to UP it is neither the case that $p \,\&\, q$ nor that $p \,\&\, \sim q$. That is, according to UP, we should find at least one row in the truth table of D3

	1	2	3	4	5	6	7
	p	q	$\sim q$	$p \,\&\, q$	$p \,\&\, \sim q$	4 v 5	$1 \rightarrow 6$
1	T	T	F	T	F	T	T
2	T	F	T	F	T	T	T
3	F	T	F	F	F	F	T
4	F	F	T	F	F	F	T

where p is true and $(p \,\&\, q)$ and $(p \,\&\, \sim q)$ are both false, i.e., where column 7 would show one F. However, inspection shows no such row in the truth table.In rows 3 and 4 where $(p \,\&\, q)$ and $(p \,\&\, \sim q)$ are both false, p is also false; in rows 1 and 2 where p is true, $(p \,\&\, q)$ v $(p \,\&\, \sim q)$ is also true even though the two disjuncts alternate truth values.

Now, D1 and D3 are tautologies if and only if standard semantics for logical connectives are assumed. Therefore, QM, which must accept UP and must reject the tautological status of D1 and D3, must also reject standard semantics for logical connectives. What does this imply for Jauch's *reductio*?

---

[15] The equivalence class of D3 includes such tautologies as $p \rightarrow (p \,\&\, (q$ v $\sim q))$ and $p \rightarrow (p \,\&\, (q \rightarrow q))$, whose status is also excluded by UP. Logical connectives being interdefinable, D1 also has an equivalence class with the same consequences.

[16] This example is adapted from David W. Cohen, *An Introduction to Hilbert Space and Quantum Logic* (New York: Springer-Verlag, 1989), 93.

6. A Problem for Jauch's *Reductio*

Jauch no doubt would have claimed that his argument for Proposition 1 valid in the intuitive sense that no errors of logic were committed, which is true as we have seen. If asked, he would have claimed further that his argument was valid in the sense of 'valid' that applies to all valid mathematical proofs. I wish to show that Jauch's acceptance of UP[17] and his implicit rejection of D1 and D3 as tautologies undermines the second claim.[18]

Because validity has a syntactic as well as a semantic meaning, proving this point requires two arguments. Let us take them in turn.

Argument 1: Syntactic Validity

(a1) Jauch's argument is valid according to the standard definition of syntactic validity only if inferred steps are derived according to standard rules of logic.

(b1) If inferred steps are derived according to standard rules of logic, then standard semantics define logical connectives in standard rules of logic applied.[19]

(c1) If standard semantics define logical connectives in standard rules of logic applied, then standard semantics define logical connectives in all rules of logic.

(d1) If standard semantics define logical connectives in all rules of logic, then they define logical connectives in D1.[20]

(e1) If standard semantics define logical connectives in D1, then D1 is a tautology.

(f1) If D1 is a tautology, there are no truth-value assignments under which both disjuncts of $(p \,\&\, q) \vee (p \,\&\, r)$ are false and $p \,\&\, (q \vee r)$ is true.

(g1) If there are no truth-value assignments under which both disjuncts of $(p \,\&\, q) \vee (p \,\&\, r)$ are false and $p \,\&\, (q \vee r)$ is true, the Uncertainty Principle is false.

(h1) The Uncertainty Principle is true.

Therefore,

(i1) Jauch's argument is not valid according to the standard definition of syntactic validity.

---

[17] Jauch, *Foundations of Quantum Mechanics*, 162.

[18] Hans Reichenbach, a proponent of three-valued logic in QM, stated that in QM "[t]he two distributive rules hold in the same form as in two-valued logic." (Hans Reichenbach, *Philosophic Foundations of Quantum Mechanics* (Berkeley and Los Angeles: University of California Press, 1944), 156.)

[19] This premise would need to be restated slightly to run the argument with D3 because D3 is not a rule of logic. There is no need to do that for present purpose.

[20] It is sufficient for present purposes to focus only on the inconsistency between UP and D1.

Argument 2: Semantic Validity

(a2) Jauch's argument is valid according to the standard definition of semantic validity only if for any truth-functional interpretation of logically compound premises and conclusion, if the premises are true, then the conclusion is true.

(b2) For any truth-functional interpretation of logically compound premises and conclusion, if the premises are true, then the conclusion is true only if standard semantics define logical connectives in premises and conclusion.

(c2) If standard semantics define logical connectives in premises and conclusion, then standard semantics define conjunction, disjunction and material implication.

(d2) If standard semantics define conjunction, disjunction and material implication, then D1 is a tautology.

(e2) If D1 is a tautology, then there are no truth-value assignments under which both disjuncts of $(p \,\&\, q) \vee (p \,\&\, r)$ are false and $p \,\&\, (q \vee r)$ is true.

(f2) If there are no truth-value assignments under which both disjuncts of $(p \,\&\, q) \vee (p \,\&\, r)$ are false and $p \,\&\, (q \vee r)$ is true, then the Uncertainty Principle is false.

(g2) The Uncertainty Principle is true.

Therefore,

(h2) Jauch's argument is not valid according to the standard definition of semantic validity.

Therefore,

(G) Jauch's argument is not valid according to standard definitions of semantic and semantic validity.

Both arguments are logically correct. But are their premises true?

7. Defense of Argument 1

Premise (a1): This follows from the definition of syntactic validity.
Premise (b1): Rules of logic applied to derive steps in the FPV of Jauch's argument are logically compound, so that standard semantics for logical connectives are automatically assumed.
Premise (c1): It cannot be the case that some rules of logic assume standard semantics for logical connectives and some do not.
Premise (d1): This premise is true because D1 is part of a rule of logic.
Premise (e1): This follows from the truth table of D1.
Premise (f1): This also follows from the truth table of D1.
Premise (g1): To see that this premise is true, consider its contrapositive:

(g1*) If the Uncertainty Principle is true, there are truth-value assignments under which both disjuncts of $(p \,\&\, q) \vee (p \,\&\, r)$ are false and $p \,\&\, (q \vee r)$ is true.

(g1*) is true based on the three propositions p, q and r specified above:

p is the proposition that the momentum of particle x is in the interval [0, +1/6],

q is the proposition that the position of particle x is in the interval [–1, +1],

r is the proposition that the position of particle x is in the interval [+1, +3]

If the Uncertainty Principle is true, then (g1*) is true given p, q and r as above. Since (g1*) and (g1) are equivalent, it follows that (g1) is also true.

Premise (h1): UP must be assumed to be true, certainly by QM theorists.

8. Defense of Argument 2

Premise (a2): This premise states a necessary condition of semantic validity.
Premise (b2): The discussion above of semantic validity justifies this premise.
Premise (c2): The truth table proving semantic validity of a portion of Jauch's argument,

	1	2	3	4	5	6
1	r	s	$\sim s$	$r \rightarrow \sim s$	$r \,\&\, (r \rightarrow \sim s)$	$(r \,\&\, (r \rightarrow \sim s)) \rightarrow \sim s$
2	T	T	F	F	F	T
3	T	F	T	T	T	T
4	F	T	F	T	F	T
5	F	F	T	T	F	T

shows that if standard semantics define logical connectives in premises (columns 1 and 4) and conclusion (column 5), then standard semantics define conjunction, disjunction and material implication. Disjunction occurs in the column 5 sentence because material implication is definable in terms of it.
Premise (d2): This follows from the truth table of D1.
Premise (e2): This is the same as premise (f1) above.
Premise (f2): This is the same as premise (g1) above.
Premise (g2): This is the same as premise (h1) above.

9. Rescuing Semantic Epistemology

It would be quite an undertaking to devise a non-standard concept of derivability – call it 'derivability*' – and then complicate SES1 as follows:

> (SES2) Where *z* is a wff of a scientific language *SL*, *z*-is-evident-in-*SL* for S =Df (i) Either there is a derivation-in-*SL* or a derivation*-in-*SL* of *z* from true-in-*SL* instrumental-accuracy-law-sentences-of-*SL* and initial-condition-sentences-of-*SL*.

This is unnecessary. We can borrow the disjunctive form of SES2 and then rely on the fact that QM makes essential use of the concept of probability:

> (SES3) Where *z* is a wff of a scientific language *SL*, *z*-is-evident-in-*SL* for S =Df Either (i) there is a derivation-in-*SL* of *z* from true-in-*SL* instrumental-accuracy-law-sentences-of-*SL* and initial-condition-sentences-of-*SL*, or (ii) the probability of *z* is certainty or practically certainty relative to true-in-*SL* instrumental-accuracy-law-sentences-of-*SL* and initial-condition-sentences-of-*SL*; and either (iii) the derivation-in-*SL* of *z* is believed-in-*SL* by S, or (iv) or the relative probability of *z* as certainty or practically certainty is believed-in-*SL* by S.

A link weaker than deduction may enable us to widen the circle of semantic knowledge to include fields not considered 'hard' sciences such as psychology, anthropology and sociology. Thus, SES3 addresses a concern raised in the earlier article.[21] The semantic definition of scientific knowledge can remain unchanged.[22]

10. Some QM Responses Considered

It is far beyond the scope of this article to evaluate efforts to cope with the unavailability of standard logic in QM by replacing it with what has come to be called 'quantum logic,'[23] including how quantum logic might formulate an FPV of Jauch's *reductio* argument. Instead, let us consider two interesting strategies that QM proponents might suggest to counter Arguments 1 and 2.

---

[21] Cusmariu, "Toward a Semantic Approach," 542. I realize more needs to be said to make clear how SES3 would cover semantic knowledge in fields not considered 'hard' sciences. However, the matter is too complex to treat adequately in an article of this scope.

[22] Cusmariu, "Toward a Semantic Approach," 536.

[23] The *locus classicus* of what came to be called 'quantum logic' is Garrett Birkhoff and John von Neumann, "The logic of quantum mechanics," *Annals of Physics* 37 (1936), 823-43. An excellent review of the issues is Peter Gibbins, *Particles and Paradoxes* (Cambridge: Cambridge University Pres, 1987). Philosophical issues in QM are addressed by contributors to *The Wave Function*, eds. Alyssa Ney and David Z. Albert (Oxford: Oxford University Press, 2013). See also Gabriel Târziu, "Quantum vs Classical Logic: The Revisionist Approach," *Logos & Episteme. An International Journal of Epistemology* III, 4 (2012): 579-590; and Pierre Uzan, "Logique Quantique et Intrication," *Logos & Episteme. An International Journal of Epistemology* V, 3 (2014): 245-263.

STRATEGY 1: Redefine logical connectives using three-valued logic.

Comment: Hans Reichenbach has proposed redefining logical connectives using three-valued logic as a way of avoiding having to characterize statements about unobserved entities as meaningless.[24] Does adding a third truth-value, *Indeterminate*, and building new truth tables for the usual logical connectives resolve the problem?[25]

It does not, for D1 as well as D3. In the D1-associated scenario, the three propositions p, q and r are all true.

> p is the proposition that the particle has momentum in the interval [0, +1/6],
>
> q is the proposition that the particle is in the interval [–1, +1], and
>
> r is the proposition that the particle is in the interval [+1, +3].

The truth value of p, q and r is not *Indeterminate*; nor are p, q and r incompatible with the Uncertainty Principle taken singly. Given that the truth value of p, q and r taken singly is *True* and not *Indeterminate*, it follows from Reichenbach's own revised truth tables[26] that truth-functional combinations of p, q and r will also not be *Indeterminate*, including material implication in D1. Moreover, if p, q and r all have the truth value True, D1 would not turn out to have the truth value *Indeterminate* even if material implication in D1 is replaced by counterparts that Reichenbach calls 'alternative implication' and 'quasi implication.'[27] Thus, Strategy 2 does not enable QM to avoid having to deny that D1 is a tautology.

STRATEGY 2: QM needs only 'the mathematics of approximation.'

Comment 1: This strategy would appeal to a distinction the mathematician Felix Klein drew in a book originally published in 1908:[28]

---

[24] Reichenbach, *Philosophic Foundations*, 144-168.

[25] Peter Gibbins writes: "There are those that try to impose on quantum mechanics a logic that does not arise naturally from the formalism of the theory. Such is Reichenbach's interpretation which employs a 3-valued truth functional logic and which is generally admitted to be a nonstarter." Gibbins, *Particles and Paradoxes*, 124. Gibbins also makes a startling admission: "… what the [logical] connectives mean is a real problem in the philosophy of quantum mechanics. All attractive routes for defining them independently of the formalism of quantum mechanics seem to be blocked (I think they are blocked.) If this is so, the scope of quantum logic as a 'logic of the world' will be restricted (as I think it is)." Gibbins, *Particles and Paradoxes*, 140.

[26] Reichenbach, *Philosophic Foundations*, 151.

[27] Reichenbach, *Philosophic Foundations*, 151.

[28] Felix Klein, *Elementary Mathematics from an Advanced Standpoint: Arithmetic, Algebra, and Analysis* (New York: Cosimo Classics, 2009), 36.

> [I]t is natural to divide mathematics into two parts, which have been called *mathematics of approximation* and the *mathematics of precision*. If we desire to explain this difference by an interpretation of the equation $f(x) = 0$, we may note that, in the mathematics of approximation one is not concerned that $f(x)$ should be *exactly* zero, but merely that its absolute value $| f(x) |$ should *remain below the attainable threshold of exactness* ε. The symbol $f(x) = 0$ is merely an abbreviation for the inequality $| f(x) | < \varepsilon$, with which one is really concerned. It is only in the mathematics of precision that one insists that the equation $f(x) = 0$ be exactly satisfied.

Comment 2: Strategy 2 suggests an instrumentalist approach to mathematics, according to which mathematical resources are merely tools for computation, measurement, approximation, and the like. In QM, such a viewpoint is sometimes expressed by the admonition to "shut up and calculate." Thus, QM physicists have argued that the theory has been confirmed by experiments, has undeniable explanatory and predictive power, and that can be the end of it as far as physics is concerned.

Well and good but instrumentalism does not imply that any of the premises of Arguments 1 and 2 are false. Indeed, how could it? Those premises belong in the realm of logic alongside the 'mathematics of precision' – just as Plato thought. In any case, Klein was surely correct to note, at the end of the passage quoted above, that the mathematics of precision provides "valuable and indeed indispensible support for the development of mathematics of approximation."[29] The claim that QM needs only the mathematics of approximation is wishful thinking.

11. Concluding Remarks

Physics from Newton to Einstein is compatible with the logic that Russell and Whitehead placed at the foundations of mathematics in *Principia Mathematica*, including the concepts of proof and validity they worked hard to clarify. If the arguments presented here are sound, such compatibility does not hold for quantum mechanics – a disconnect that seems radical enough to warrant considering the theory a paradigm shift and may help explain why Richard Feynman famously quipped[30] "I think I can safely say that nobody understands quantum mechanics."[31]

---

[29] Klein, *Elementary Mathematics*, 36.

[30] Richard Feynman, *The Character of Physical Law: The Messenger Lectures* (New York: Modern Library, 1994), 129.

[31] Takis Hartonas and Gary Rosenkrantz provided helpful comments on earlier versions of this article.

A PROPOS DU RENOUVEAU ANNONCE DE LA METAPHYSIQUE

Pierre UZAN

ABSTRACT: In this paper, we evaluate the project of resurgence of metaphysics based on the pecularity of the quantum domain, a project that is supported by some contemporary philosophers. Beyond the general arguments against scientific realism that are still applicable here, we show that this project is faced with the three following issues that, we believe, make it unrealizable: (a) the problem raised by the realistic interpretation of the wave function, as a description of a 'concrete physical fact' of the independent reality; (b) the lack of any experimental counterpart of the (non-local) hidden variables quantum theories, and, in some cases, their incompatibility with the quantum predictions; and (c) the fact that the key-properties of quantum phenomena, like their non-locality, essentially depend on the observables that are used for their description and cannot then be assigned to any 'independent' reality.

KEYWORDS: entanglement, hidden variables theories, interpretation of the wave function, metaphysics of science, quantum physics

1. Introduction

La « renaissance de la métaphysique » fondée sur « les connaissances qu'apportent les sciences » que constate Michael Esfeld[1] est bien illustrée par les interprétations à visée ontologique de la mécanique quantique. Nous pouvons, en particulier, mentionner la « mécanique Bohmienne »[2] qui reprend et prolonge la théorie de l'onde-pilote de Louis de Broglie, l'interprétation réaliste du processus de réduction de la fonction d'onde en terme de localisation spontanée proposée par Girhardi, Rimini et Weber,[3] qui sera notée ci-dessous « théorie GRW », ou la philosophie du « réel voilé » développée par Bernard d'Espagnat à partir des fondements de la mécanique quantique.[4] Plus récemment, Esfeld[5] et Dorato[6] ont

[1] Michael Esfeld, « La philosophie comme métaphysique des sciences », *Studia Philosophica* 66 (2007): 61-76.

[2] David Bohm and Basil Hiley, *The Undivided Universe* (London: Routledge, 1993).

[3] GianCarlo Girhardi, Alberto Rimini and Tullio Weber "Unified dynamics for microscopic and macroscopic systems," *Physical Review* D, 34, 2 (1986): 470-491.

[4] Bernard d'Espagnat, *Le Réel Voilé* (Paris : Fayard, 1994).

[5] Michael Esfeld, "How to Account for Quantum Non-locality: Ontic Structural Realism and the Primitive Ontology of Quantum Physics," *Synthèse*, online 11th Septembre 2014.

proposé la construction de ce que l'on pourrait appeler une « métaphysique des relations » qui combine le réalisme structural ontologique introduit par Ladyman[7] et les interprétations ontologiques de la mécanique quantique mentionnées ci-dessus – mécanique Bohminenne ou théorie GRW. Mentionnons enfin la proposition de Dorato[8] de considérer l'intrication, qui est une propriété caractéristique du domaine quantique, comme le fondement du « nouvel ordre fondamental de la *nature* ».[9]

Cependant, malgré la cohérence de ces propositions à visée ontologique, pouvons-nous accepter que ce que nous apprend la science (ici, la physique quantique) contribue vraiment à l'élaboration d'une « une vision cohérente et complète du monde »[10] si ce « monde » est, comme c'est le cas dans ces propositions, conçu comme réel-en-soi, indépendant de nos moyens d'investigations et de nos représentations ? Pouvons-nous souscrire à l'idée selon laquelle la science aurait pour fonction et pour but de décrire la réalité indépendante, une idée souvent admise sans discussion par la plupart des scientifiques ? Ou devons-nous plutôt faire preuve de prudence en maintenant la distinction kantienne entre phénomène et noumène, entre la construction du phénomène qui fait appel aux formes a priori de la connaissance imposées par le sujet et l'existence d'une hypothétique réalité indépendante auquelle la science n'aurait pas accès ?

Cette question traditionnelle de la métaphysique a été de nombreuses fois discutée (par exemple par Boyd[11] ou Chakravartty[12]) et sera ici reconsidérée en nous référant tout particulièrement aux interprétations ontologiques de la mécanique quantique mentionnées ci-dessus. A la section 2, nous commencerons par rappeler quelques arguments généraux qui ont été formulés à l'encontre du réalisme scientifique. En particulier, nous insisterons sur un argument constructiviste, issu de la philosophie critique, qui peut être considéré comme

---

[6] Mauro Dorato and Michael Esfeld, "GRW as an ontology of dispositions," *Studies in History and Philosophy of Science, Part B,* 41, 1 (2010): 41-49.

[7] James Ladyman, "What is Structural Realism?" *Studies in History and Philosophy of Science*, 29 (1998): 409–424.

[8] Mauro Dorato, "Laws of Nature and the Reality of the Wave Function," *Synthese (2016),* forthcoming.

[9] Nous avons souligné le mot « nature » qui témoigne de l'engagement ontologique de l'auteur.

[10] Michael Esfeld, « La philosophie comme métaphysique des sciences ».

[11] Richard N. Boyd, "On the Current Status of the Issue of Scientific Realism," *Erkenntnis* 19, (1983): 445-90.

[12] Anjan Chakravartty, *A Metaphysics for Scientific Realism: Knowing the Unobservable* (Cambridge: Cambridge University Press, 2007).

l'argument « positif » majeur en ce qu'il montre *comment se constitue le monde des objets*, mettant ainsi en difficulté l'idée que les théories scientifiques décrivent le monde « tel qu'il est ». Nous soutiendrons cet argument en nous référant aux travaux de Duhem et de Quine et, plus généralement, à la philosophie kantienne et, dans son prolongement, aux analyses que fait Cassirer du processus de la connaissance. Dans la section 3, nous montrerons qu'en plus des arguments généraux qui ont été proposé à l'encontre du réalisme scientifique, des arguments spécifiques à l'encontre des interprétations ontologiques de la mécanique quantique peuvent être formulés. Ces derniers arguments, d'ordre plus technique, se rapportent aux trois thématiques suivantes qui seront traitées successivement : (a) la question du statut de la fonction d'onde (et, plus généralement, de l'entité formelle représentant l'état d'un système dans la théorie quantique); (b) la question de la complétude de la théorie quantique et de la possibilité d'y ajouter des variables supplémentaires dans le but de décrire les mécanismes sous-jacents des phénomènes ; et (c) la question de la légitimité qu'il y aurait à attribuer à la réalité indépendante certains traits caractéristiques du domaine quantique, comme en particulier la non-localité qui est considérée dans ces approches comme une caractéristique structurelle intrinsèque de la réalité indépendante.

2. Le processus de constitution de l'objectivité

La science nous renseigne-t-elle sur la réalité indépendante ? Cette question, qui constitue la problématique essentielle de la philosophie des sciences, peut être reposée de la façon suivante qui fait entrevoir son aspect paradoxal: Les lois scientifiques, qui sont toujours formulées dans le cadre d'un certain paradigme théorico-expérimental, peuvent-elles être considérées comme celles régissant une réalité indépendante de nos moyens d'investigations et de nos représentations ?

Plusieurs arguments généraux ont été proposés à l'encontre du réalisme scientifique (voir la synthèse proposée dans le livre de Chakravartty cité ci-dessus). Par exemple, un argument extrêmement puissant est celui de la sous-détermination des théories par l'expérience (ou thèse de Quine-Duhem) d'où sont directement dérivés des arguments qui remettent en question l'idée que la science s'approche de plus en plus de la vérité, voire la notion même de « vérité » scientifique. L'argument de la sous-détermination des théories par l'expérience mentionne le fait que des théories différentes peuvent décrire les mêmes phénomènes. Ce fut par exemple le cas de la théorie corpusculaire de la lumière formulée par Newton et de la théorie ondulatoire de la lumière proposée par Huygens et développée par Young et par Fresnel puisque ces deux théories rivales pouvaient expliquer la propagation rectiligne de la lumière ainsi que les

phénomènes de réfraction. L'existence de telles théories « empiriquement équivalentes » qui expliquent ces phénomènes en se référant pourtant à des entités et des mécanismes *différents* (dans l'exemple considéré : corpuscules pour l'une, ondes pour l'autre, qui sont régis par leurs propres lois) met ainsi en difficulté l'idée qu'une théorie serait plus « vraie » qu'une autre, qu'elle décrirait plus fidèlement les entités de la réalité indépendante.

Cependant, nous nous concentrerons ici sur un seul argument que l'on pourrait qualifier de « positif » dans la mesure où il montre *comment* se construit l'objectivité de la science et supplante donc tous les autres arguments dont le but est d'essayer de mettre en défaut le réalisme en montrant ses faiblesses – ces derniers arguments peuvent donc être considérés comme des arguments critiques « négatifs ». L'idée-force capable, sinon de remettre en question, de montrer tout du moins le caractère purement spéculatif de tout tentative d'interprétation ontologique des théories scientifiques est le constat que l'objectivité scientifique résulte d'un *processus d'élaboration sémantique* effectué dans un contexte scientifique et, plus généralement, socio-culturel donné – ce qui introduit, en outre, une dimension historique à la notion de « vérité scientifique ». En effet, contrairement à ce que prétend une vision naïve de la science, qui est souvent acceptée sans discussion dans les milieux scientifiques, les « objets » dont nous parle la science (électrons, molécules, gènes,…) ne sont donc pas simplement « découverts » lors d'observations passives de « faits » bruts mais ils sont *constitués.* L'idée de constitution de l'objectivité scientifique a été formulée de façon concise par Duhem :[13]

> [le phénomène qu'enregistre un expérimentateur est] une interprétation qui substitue aux données concrètes réellement recueillies par l'observation des représentations abstraites et symboliques qui leur correspondent en vertu des théories admises par l'observateur.

Plusieurs illustrations de ce processus de constitution de l'objectivité scientifique ont été proposés dans la littérature, comme c'est le cas du processus complexe d'élaboration du concept de « photon » qui a été clairement analysé par Léna Soler.[14] Prenons ici un exemple plus simple mais néanmoins très significatif, celui de la « découverte » de la radioactivité par Becquerel, que l'on présente souvent comme un fait brut détaché de son contexte théorique. Cette « découverte » est, en fait, le résultat d'une élaboration sémantique complexe à

[13] Pierre Duhem, *La théorie physique, son objet, sa structure,* chap IV, §1. (Paris : Vrin, 1981), 209.

[14] Léna Soler, *L'emergence d'un nouvel objet symbolique* : *le photon*. Thèse de Doctorat sous la direction de Michel Bitbol, Université Paris 1 (1997).

partir de la perception de traces – que l'on suppose, pour simplifier, constituer les ultimes « données expérimentales » – laissées par des sels d'uranium sur des émulsions photographiques et de l'ensemble des théories physiques de l'époque – et en particulier ici, la théorie atomique. Il fallait, en effet, avoir au préalable élaboré le concept d'atome et, en outre, d'un atome composé de particules encore plus élémentaires pour analyser ces traces comme le résultat de l'impact de particules émises spontanément par les atomes d'uranium.

Un « fait » scientifique est donc toujours le résultat d'un processus d'interprétation dans le cadre du paradigme scientifique en vigueur, d'où l'extrême difficulté, voire l'impossibilité, de distinguer entre la part de notre savoir qui relèverait d'un hypothétique monde-en-soi et de celle relevant de notre façon d'organiser nos connaissances. Ce point a été illustré par Quine dans ce passage bien connu :[15]

> La tradition de nos pères est un tissu de phrases. Entre nos mains, il se développe et change, par ses révisions et additions délibérées plus ou moins arbitraires et qui lui sont propres, plus ou moins directement occasionnées par la stimulation. C'est une tradition couleur gris pâle, noire de fait et blanche de convention. Mais je n'ai trouvé aucune raisonmajeure de conclure qu'il s'y trouve le moindre fil tout à fait noir, ou tout à fait blanc.

L'idée de constitution de l'objectivité, qui s'oppose à celle de découverte d'un monde-en-soi déjà constitué, a été analysée de façon rigoureuse par Kant et les philosophes post-kantiens. La connaissance d'un objet est, selon Kant, une mise en forme, imposée par le sujet de la connaissance, d'une « matière » donnée. La connaissance d'un objet fait donc intervenir deux sortes d'éléments : ceux qui dépendent du sujet et qui constituent la forme a priori de la connaissance et ceux dépendant de l'objet lui-même et qui en constituent la « matière ». Les formes a priori de la connaissance seraient ainsi, selon cette conception, les cadre universels et nécessaires à travers lesquels l'esprit humain appréhende le monde : les formes a priori de la sensibilité (l'espace et le temps), qui constituent notre façon de le percevoir, et les formes a priori de l'entendement, catégories ou concepts purs (comme celles de causalité ou de substance), qui constituent notre façon de le concevoir.

Il est cependant possible d'affiner le modèle kantien de la connaissance où l'identification de l'a priori et de l'inné conduit à une conception statique et dualiste de la connaissance : d'un côté les formes a priori intemporelles et de l'autre les « données » contingentes, la matière de la connaissance. Cassirer a, en

[15] Willard V. O. Quine, "Carnap and Logical Truth," in *The Philosophy of Rudolf Carnap,* ed. Paul A. Schlipp (La Salle: Open Court, 1963), 374.

effet, proposé de re-définir le processus de la connaissance en terme de mise en forme *symbolique*, comme un processus *d'élaboration de la signification* qui constituerait, de fait, notre seul mode d'objectification.[16] Selon Cassirer, toutes nos théories[17] sur le monde, toutes les informations (d'origine expérimentale ou non) dont nous disposons, mais aussi toutes les actions entreprises dans le but de l'observer constituent, respectivement, en tant que *formes de représentation, données interprétées dans ce contexte théorique et procédures d'acquisition* de ces données, autant d'éléments *définissant* le processus de la connaissance. Les formes symboliques, que Cassirer classe en langage, pensée mythique, artistique, religieuse et scientifique sont « les modalités de donation du sens » et constituent, selon lui, « la trame enchevêtrée de l'expérience humaine que tout progrès dans la pensée et l'expérience de l'Homme complique et renforce ». Toute réalité étant médiatisée par des formes linguistiques, artistiques, scientifiques, par des symboles mythiques ou des rites religieux, nous pouvons ainsi dire que *l'Homme vit dans un Univers symbolique*: il n'a accès qu'à un monde de représentations et jamais à un monde-en-soi qui lui serait donné. L'univers symbolique ou univers de représentation de l'Homme résulte donc d'un processus d'élaboration sémantique et constitue toute sa réalité, réalité qui est en fait délimitée par son langage et l'ensemble de toutes les formes culturelles à travers lesquelles et par lesquelles se constitue pour lui toute « objectivité ».

Cette analyse du processus de la connaissance ainsi que les remarques précédentes de Duhem et de Quine militent donc fortement en faveur de l'idée que *l'objectivité se construit*, *qu'elle résulte d'un processus d'élaboration sémantique*, ce qui laisse ainsi peu de place au projet de développement d'une métaphysique des sciences. En outre, nos moyens d'investigation étant tributaires d'un paradigme donné (et en général transitoire), qu'est-ce qui nous autoriserait à affirmer que la science d'aujourd'hui, plus que celle d'hier ou de demain qui peuvent relever de paradigmes différents, voire incommensurables,[18] nous révèlerait les « véritables » structures de la réalité indépendante ? Par exemple, la réalité indépendante serait-elle constituée d'une substance unique (Thalès), d'atomes (Lucrèce, Démocrite, Dalton), d'éther (Descartes, Fresnel, Maxwell) ou,

[16] Ernst Cassirer, *La philosophie des formes symboliques* (Paris : Editions de minuit, 1927), 43.

[17] « Théorie » au sens le plus général d'ensemble d'idées et de concepts de tout ordre plus ou moins structurés contribuant à notre vision du monde.

[18] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962). Trad. Française : *La structure des révolutions scientifiques* (Paris: Flamarion, coll. Champs, 1983).

en référence aux théories cosmologiques récentes, n'est-il qu'une fluctuation du vide quantique ?

3. Trois arguments spécifiques à l'encontre des interprétations à visée ontologique de la mécanique quantique

Si les arguments généraux brièvement rapportés ci-dessus à l'encontre du réalisme scientifique s'appliquent bien sûr aux interprétations ontologiques de la mécanique quantique, il est, en outre, possible de formuler des arguments spécifiques à l'encontre de ces interprétations. La mécanique quantique a, en effet, introduit des concepts nouveaux qui sont en rupture avec ceux des mécaniques classiques et relativistes et a développé un formalisme permettant de représenter rigoureusement ces concepts. Afin de formuler ces arguments spécifiques, il nous faut commencer par rappeler brièvement les concepts typiquement quantiques qui ont poussé certains à clamer le renouveau de la métaphysique. Nous montrerons, en particulier, en quoi ces derniers concepts se distinguent de ceux des mécaniques classique et relativiste.[19]

Selon la mécanique quantique standard, dont les principes sont brièvement rappelés ici sans interprétation particulière,[20] l'état d'un système n'est pas représenté, comme en mécanique classique, par un point de l'espace des phases, c'est à dire par la donnée de sa position et de sa vitesse, mais par un vecteur de l'espace vectoriel des états bâti sur le corps des nombres complexes (espace de Hilbert). Cet état dont l'évolution temporelle est donnée par l'équation de Schrödinger permet le calcul des prédictions via la règle de Born.[21] Corrélativement, les grandeurs physiques que nous pouvons mesurer sur ce système ne peuvent plus s'écrire comme des fonctions des coordonnées de ce point-état mais comme des opérateurs auto-adjoints, des « observables », agissant sur l'espace des états. Une première caractéristique essentielle de la mécanique

[19] Pour un exposé plus complet de la théorie quantique standard, le lecteur pourra par exemple se reporter au manuel de *Mécanique Quantique* de Claude Cohen Tanoudji, Bernard Diu et Franck Laloë (Paris : Hermann, 1977).

[20] Ce qui veut dire que seule la structure mathématique minimale permettant de décrire les phénoménes quantiques, celle qui est présentée dans les manuels universitaires, sera ici utilisée. La question de l'interprétation de ce formalisme qui est un enjeu essentiel dans cet article sera analysée ci-après.

[21] La règle de Born peut en fait être considérée comme une utilisation, dans le domaine quantique, de la seule mesure de probabilités qu'il soit est possible de définir sur l'ensemble des sous-espaces clos d'un espace de Hilbert. Ce dernier point a été montré par Andrew M. Gleason, "Measures on the Closed Subspaces of a Hilbert Space," *Journal of Mathematics and Mechanics,* 6 (1957): 885-893.

quantique est que, contrairement aux prédictions déterministes de la mécanique newtonienne où la connaissance de l'état initial d'un système permet de déterminer avec certitude son état ultérieur, la mécanique quantique est une théorie fondamentalement probabiliste[22] qui ne fournit (par la règle de Born) qu'une distribution de probabilités de mesurer une certaine valeur pour une observable donnée – le spectre de valeurs possibles de cette observable s'identifiant ainsi à celui de ces valeurs propres. Outre la représentation qui est faite de l'état d'un système par un vecteur ou une « fonction d'onde »[23] et le caractère probabiliste irréductible de la théorie quantique, cette dernière a des propriétés bien particulières que ni la mécanique classique ni la mécanique relativiste ne partagent. D'une part, la théorie quantique est une théorie *contextuelle*,[24] ce qui veut dire que la distribution de probabilité associée à la mesure d'une observable dépend de *l'ensemble du dispositif expérimental* (et, en particulier, des autres observables que l'on décide de mesurer conjointement). D'autre part, c'est une théorie *non-locale*, ce qui signifie que la distribution de probabilités associée à une observable mesurée en une région R de l'espace-temps[25] peut dépendre, comme l'a confirmé la violation expérimentale des inégalités de Bell (voir, par exemple, l'article de Clauser *et al.*),[26] de celles associées à la mesure d'autres observables mesurées dans des régions spatialement séparées[27] de R, ce qui est en particulier le cas lorsque ces mesures sont effectuées sur les parties d'un système préparé dans un état intriqué. La propriété d' *« intrication »* de l'état d'un système composé, qui peut être considérée comme la caractéristique la plus

[22] Ce qui veut dire que les probabilités calculées par la règle de Born ne peuvent s'interpréter comme des probabilités d'ignorance, comme en mécanique statistique classique. Leur existence est directement liée au formalisme des observables utilisé (qui ne se réduisent pas à des fonctions calculant des valeurs uniques pour les grandeurs mesurées).

[23]La fonction d'onde est une description de l'état d'un système selon un mode de représentation particulier (en « position », historiquement). Elle est obtenue dans ce cas en projetant le vecteur d'état sur une base des vecteurs propres de la position.

[24] La contextualité de la théorie quantique a été montrée par Simon B. Kochen et Ernst Specker – voir "The Problem of Hidden Variables in Quantum Mechanics, " *Journal of Mathematics and Mechanics* 17, (1967) : 59–87.

[25] Ou, autrement dit, en un point de l'espace et à un instant donné.

[26] John F. Clauser, Michael Horne, Abner Shimony and Richard A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," *Physical Review Letters* 23, (1969): 880-884.

[27] Ce qui veut dire que selon la relativité aucun signal physique ne peut relier ces deux évènements de mesure.

fondamentale de la mécanique quantique, selon les mots de Schrödinger,[28] signifie que cet état ne peut être décrit par une simple adjonction des descriptions relatives à ses sous-systèmes. L'intrication d'un état s'exprime formellement par le fait qu'il ne peut être écrit comme le produit d'états de ses sous-systèmes, ce qui a pour conséquence que des corrélations *non-locales*, ne pouvant être expliquées « classiquement »,[29] peuvent ainsi être observées entre les observables définies sur ces sous-systèmes.

Les interprétations de la mécanique quantique qui ont été proposées se réfèrent à au moins l'une des trois thématiques suivantes qui ne sont bien sûr pas indépendantes mais que, pour des raisons didactiques, nous allons considérer successivement. Il s'agit de (a) la question relative au statut de l'entité représentant l'état d'un système dans la théorie quantique, le vecteur d'état ou la « fonction d'onde » ; (b) la question de la complétude de la théorie quantique, à laquelle certains répondent par l'ajout des variables supplémentaires qui décriraient les hypothétiques mécanismes sous-jacents aux phénomènes, et (c) celle de la légitimité qu'il y aurait à attribuer certaines propriétés typiques des phénomènes quantiques (celle de non-localité, en particulier) à la réalité indépendante.

Question (a). La question du statut de la fonction d'onde (ou du vecteur d'état) peut se formuler de la façon suivante : la fonction d'onde donne-t-elle une description, même partielle, de la réalité indépendante ou ne doit-elle être considérée *que* comme un outil formel encodant notre connaissance de la situation expérimentale étudiée et permettant le calcul des prédictions ?

Cette question était déjà débattue par les pères fondateurs de la mécanique quantique : Schrödinger[30] soutenait, du moins au début (voir ci-après), que la fonction d'onde associée à l'état d'un système était une *onde physique* et que la réalité indépendante était composée de telles ondes physiques dont l'interaction permettrait d'expliquer l'existence de corpuscules. Aux antipodes de cette position, Bohr[31] soutenait, dans le cadre de son interprétation holistique du phénomène, que l'entité formelle représentant l'état d'un système (comme sa fonction d'onde, en particulier) *ne peut se rapporter à ses propriétés intrinsèques* mais qu'il doit être conçu comme un simple outil formel encodant les caractéristiques de

[28] Erwin Schrödinger, "Discussion of Probability Relations between Separated Systems," *Proceedings of the Cambridge Philosophical Society,* 31 (1935): 555–563 and 32 (1936): 446–451.

[29] C'est à dire par une action de proche en proche dans l'espace, que ce soit par l'existence d'une interaction causale directe ou par celle d'une cause commune.

[30] Schrödinger, "Discussion of Probability."

[31] Niels Bohr, "Foundations of Quantum Physics I," *Collected Works (1933-1958)*, Vol. 7, ed. Kalckar Jørgen (Amsterdam: Elsevier, 2008).

l'ensemble du dispositif expérimental et permettant ainsi de calculer les résultats possibles d'une mesure. L'interprétation holistique de Bohr a été précisée par Bitbol[32] lorsqu'il montre, en se référant aux analyses de J.L. Destouches[33] et de P. Destouches-Fevrier[34] relatives aux théories de la prévision, que le vecteur d'état constitue un outil de prévision probabiliste particulièrement simple « d'évènements définis non pas dans l'absolu mais *relativement à un contexte expérimental ».* En outre, cet auteur met en évidence (pp. 157-160) le rôle d' « invariant du système de coordonnées probabilistes entre *toutes* les situations expérimentales accessibles à la suite d'une préparation donnée » du vecteur d'état– c'est à dire que ce dernier est une entité *universelle* vis à vis des mesures pouvant être effectuées par suite de cette préparation. Quant à Einstein, il soutenait l'idée, avec Podolski et Rosen que la fonction d'onde nous donnait une description incomplète de la réalité et qu'afin de sauver le principe de causalité locale[35] sur lequel se fonde la relativité il fallait adjoindre à la théorie quantique des paramètres supplémentaires, permettant la description des processus « réels » sous-jacents :[36]

> While we have thus shown that the wavefunction does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible.

Schrödinger a lui-même trouvé un argument puissant à l'encontre de l'interprétation réaliste de la « fonction d'onde » comme onde physique. Cet argument tient au fait que la fonction d'onde représentative de deux systèmes en interaction (comme, en particulier, le système observé et l'appareil de mesure) ne pouvait se réduire à la simple interaction de deux ondes mais nécessitait d'introduire une nouvelle fonction d'onde pour le système composé, ce dernier devant alors être considéré comme un tout indivisible – ce qui rejoint la conception Bohrienne rapportée ci-dessus. Cet argument lui a fait dire que, finalement, bien que la fonction d'onde donne une description « complète et continue dans l'espace et le temps, sans omission ni lacune, conforme à l'idéal

[32] Michel Bitbol, *Mécanique quantique, §. 2.2.2.* (Paris : Flammarion, 1996).

[33] Jean-Louis Destouches, *Mécanique ondulatoire* (Paris : Presse Universitaire de France 6ième édition, 1971).

[34] Paulette Destouches-Février, *La structure des théories physiques* (Paris: Presses Universitaires de France, 1951).

[35]Selon ce principe, les évènements se produisant dans une région de l'espace-temps ne peuvent dépendre ou être influencés par ceux se produisant dans une région spatialement séparée.

[36] Albert Einstein, Boris Podolski and Nathan Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review 47 (*1935): 780.

classique », cette dernière ne décrit pas « ce que la nature (la matière, le rayonnement, etc…) *est* réellement. En fait, nous utilisons cette description (la description dite ondulatoire) en sachant parfaitement bien qu'elle ne correspond à *aucun* de ces termes. » [37]

S'il est aujourd'hui impossible de soutenir une interprétation réaliste naïve de la fonction d'onde, comme onde physique (au même titre qu'une onde mécanique ou électromagnétique, par exemple), il nous faut tout de même mentionner la proposition réaliste originale formulée récemment par Michael Esfeld.[38] Selon cet auteur, en plus de son rôle nomologique comme outil mathématique intervenant dans les équations d'évolution et permettant le calcul des prédictions, la fonction d'onde jouerait aussi un rôle descriptif non pas des entités qui composent un système, qui sont conçus comme des « éléments de l'ontologie primitive » (en particulier, de la mécanique Bohmienne[39]) ou de leurs propriétés intrinsèques, mais *de leur relations et de l'enchevêtrement de leurs évolutions temporelles.* Ce qui lui fait dire que la fonction d'onde encoderait *« un fait physique concret » de la réalité indépendante* qui pourrait ainsi « expliquer » le phénomène de non-localité. Cependant, s'il est incontestable que les phénomènes de non-localité trouvent leur expression formelle dans la structure mathématique de la fonction d'onde, à savoir dans les termes de corrélation entre les observables des parties du système considéré, est-il légitime de supposer que ces corrélations existent *dans la réalité indépendante* ? La légitimité de l'interprétation ontologique de la fonction d'onde repose donc ici sur celle qu'il y aurait à attribuer à la réalité indépendante les propriétés typiques des phénomènes quantiques. Cette dernière question sera abordée ci-après (question (c)) et trouvera une réponse négative, ce qui met donc en difficulté cette interprétation réaliste d'inspiration structuraliste de la fonction d'onde soutenue par Esfeld. En attendant, poursuivons notre analyse du statut de la fonction d'onde en présentant cette fois-ci un argument « positif » montrant comme *se construit* cet outil prédictif.

Spekkens[40] a récemment proposé une construction *purement épistémique* de l'entité formelle permettant le calcul des prédictions qui retrouve les propriétés essentielles du domaine quantique, telles que la non-commutativité des mesures et l'existence d'état intriqués donnant lieu aux phénomènes de non-localité. Ce

[37] Erwin Schrödinger, « *Science et humanisme* », dans Erwin Schrödinger, *Physique quantique et représentation du monde* (Paris : Seuil, 1992), 60.

[38] Esfeld, "How to Account for Quantum Non-locality."

[39] Voir ci-après (question (b)) pour un bref exposé de cette approche.

[40] Robert W. Spekkens, "In Defense of the Epistemic View of Quantum States: a Toy Theory," *Physical Review A* 75, 032110 (2007), quant-ph/0401052.

modèle repose sur un principe unique portant sur la quantité maximale d'*information* dont nous pouvons disposer sur un système. Plus précisément, ce principe affirme que si nous disposons de cette quantité d'information maximale, elle doit être égale à celle qui nous manque pour déterminer complètement l'état de ce système. Ce n'est pas notre propos d'entrer ici dans les débats relatifs à la justification de ce dernier principe (voir, par exemple, la discussion présentée par Avin[41]), mais nous pouvons voir dans les travaux de Spekkens un argument « positif » à l'encontre du réalisme de la fonction d'onde puisqu'ils montrent comment cet outil prédictif contextuel peut être construit et comment peuvent être retrouvées les propriétés fondamentales du domaine quantique *à partir de considérations purement épistémiques, c'est à dire sans faire appel à l'existence fortement objective d'« éléments de réalité » sous-jacents à ces phénomènes.*

Question (b). La question de la complétude de la théorie quantique, qui a aussi donné lieu à des débats animés entre les pères fondateurs de la mécanique quantique,[42] est une thématique essentielle de certaines approches ontologiques puisqu'elle est directement liée à la possibilité de compléter cette théorie par des variables supplémentaires (ou « cachées » car non explicitées dans la théorie quantique standard) permettant de décrire les mécanismes « réels » sous-jacents des phénomènes quantiques. Le paradigme d'une telle approche est la mécanique Bohmienne qui postule l'existence de particules réelles se comportant comme des corpuscules classiques dotés de propriétés intrinsèques (possédant, en particulier, une position et une vitesse déterminées à chaque instant) et qui seraient soumis à des processus dynamiques déterministes cachés. Ces derniers, qui permettent effectivement d'expliquer de cette façon les phénomènes quantiques (non relativistes, du moins), comme c'est le cas pour le phénomène d'interférence,[43] sont régis par un « potentiel quantique » agissant instantanément en plusieurs endroits de l'espace pour guider les particules. Cependant, comme il a été noté, le problème est qu'une telle théorie à variables cachées non-locales ne donne lieu à *aucune prédiction nouvelle* par rapport à la théorie standard.[44] En effet, il est facile de constater que la partie purement prédictive de cette théorie est strictement

[41] Shahar Avin, *A Philosopher's View of the Epistemic Interpretation of Quantum Mechanics,* Dissertation for the Master Level course, Cambridge University, 2010.

[42]Voir Einstein, Podolski and Rosen, "Can Quantum-Mechanical Description," cité ci-dessus, ainsi que Niels Bohr, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review* 48 (1935): 696-702.

[43] Chris Philippidis, Chris Dewdney and Basil J. Hiley, "Quantum Interference and the Quantum Potential," *Nuevo Cimento* 52 B 1 (1979): 15-28.

[44] Voir Bitbol, *Mécanique quantique,* cité ci-dessus, ainsi que Frederik J. Belinfante, *A Survey of Hidden Variable Theories*, § 2.16 (Oxford: Pergamon Press, 1973).

équivalente à celle de la théorie quantique standard dans la mesure où aucune outil formel prédictif nouveau ni aucune procédure prédictive additionnelle n'y sont introduits.[45] On pourrait penser que cette conclusion est liée à l'état de développement de cette théorie particulière, mais il semble que la raison soit plus profonde et s'applique à toute théorie quantique à variables « cachées » qu'il est possible de construire (c'est à dire à variables cachées non-locales ou contextuelles). Cette raison se trouve dans le théorème de Kochen et Specker qui nous dit que toute théorie reproduisant les prédictions de la mécanique quantique doit être *contextuelle* (voir ci-dessus). Ce qui a pour conséquence que dans toute théorie à variables supplémentaires reproduisant les prédictions de la mécanique quantique, comme c'est le cas de la mécanique Bohmienne, *ces variables supplémentaires sont nécessairement contextuelles*, c'est à dire que leurs valeurs dépendent de l'ensemble de la situation expérimentale. Plus précisément, la propriété de contextualité impose que ces dernières dépendent non seulement des observables que l'on choisit de mesurer mais, plus généralement, de l'entité formelle (fonction d'onde ou vecteur d'état) qui encode l'information maximale dont nous disposons sur la situation expérimentale considérée et qui permet le calcul des prédictions. D'ailleurs, cette conséquence se vérifie facilement pour la mécanique Bohmienne puisque les supposées propriétés intrinsèques des particules sont, à chaque instant, calculées *à partir de la fonction d'onde* : l'impulsion d'une particule est définie à chaque instant comme le gradient de *la phase de la fonction d'onde* alors que sa position est calculée à partir du potentiel quantique qui est lui-même défini à partir de *l'amplitude de la fonction d'onde*. Ce qui signifie que dans la mécanique Bohmienne *l'ajout de variables « supplémentaires » définies à partir de la fonction d'onde globale du système ne peut donner lieu à des prédictions qui ne pourraient être données par la théorie quantique standard, et n'a donc aucun répondant expérimental.* Par conséquent, si cette approche déterministe à visée ontologique permet à certains de mieux « comprendre » les phénomènes quantiques en faisant appel à des modèles spatiaux, aucune expérience ne pourra, par principe, confirmer ou infirmer l'existence des processus déterministes sous-jacents qui y sont supposés – ce qui rend ces approches ontologiques purement spéculatives car dénuées principiellement de toute répondant expérimental.

Un jugement plus sévère encore à l'encontre des théories à variables supplémentaires peut être formulé en prenant en compte des avancées théoriques et expérimentales récentes montrant qu'une classe importante de ces théories fondée sur des hypothèses assez raisonnables sont *incompatibles* avec la

[45] Voir Pierre Uzan, *Vers une logique du temps sémantique* (Lille: Presses Universitaires du Septentrion, 1998), 298-410.

mécanique quantique. Des tests expérimentaux falsifiant une classe assez étendue de modèles quantiques à variables cachées non-locales portant sur des mesures de l'état de polarisation de photons (effectuées par Gröblacher et son équipe[46]) ou de leur moment angulaire orbital (effectuées par Romero et son équipe[47]) ont, en effet, été rapporté récemment. Ces modèles montrent, chacun dans leur domaine, la violation théorique (par la théorie quantique standard) et expérimentale d'inégalités qui ont été démontrées par Leggett[48] et que ces auteurs adaptent à leur modèles. Les inégalités de Leggett sont obtenues en partant de conditions très générales que doivent vérifier les phénomènes d'émission et la détection de photons et, point essentiel, en relaxant l'hypothèse d'indépendance des résultats de la mesure d'une observable à l'égard des paramètres, c'est à dire vis à vis des autres observables que l'on décide de mesurer sur ce système (définies, par exemple, par la direction des polariseurs utilisés) – c'est l'élimination de cette dernière hypothèse qui introduit la non-localité pour les modèles à variable cachées considérés. Leggett avait montré que ces inégalités très générales contredisaient les prédictions de la mécanique quantique et, plus concrètement, Gröblacher et Romero ont reformulé ces inégalités pour les modèles particuliers qu'ils proposent et ont montré qu'elles sont falsifiées par l'expérience. La viabilité des théories quantiques à variables cachées non-locales est donc sérieusement remise en question, que ce soit pour leur manque de répondant expérimental (mécanique Bohmienne) ou, plus radicalement, pour leur incompatibilité avec les prédictions de la mécanique quantique standard (qui sont, elles, bien vérifiées).

Question (c). Il semble que la légitimité d'attribuer à la réalité indépendante certaines propriétés typiques des phénomènes quantiques n'ait pas été encore bien explorée, ce qui donne lieu à des sur-interprétations injustifiées des concepts quantiques que nous nous proposons de dénoncer ici. Cette critique pourra s'appliquer à trois propositions qui se fondent sur l'idée que la réalité indépendante devrait partager certaines propriétés du domaine quantique : il s'agit dela conception du « réel voilé » soutenue par Bernard d'Espagnat et de propositions plus récentes, formulées par Michael Esfeld et par Mauro Dorato, d'une interprétation ontologique de la non-localité.

[46] Simon Gröblacher et al.,"An Experimental Test of Non-local Realism," *Nature* 446 (2007): 871-875.

[47] Jacquiline Romero et al., "Violation of Leggett Inequalities in Orbital Angular Momentum Subspaces," *New Journal of Physics* Vol. 12 (12), 123007 (2010).

[48] Anthony Leggett, "Nonlocal Hidden-Variable Theories and Quantum Mechanics: An Incompatibility Theorem," *Foundations of Physics* 33 (2003): 1469-1493.

Dans son livre *Le Réel Voilé*, Bernard d'Espagnat[49] reconnaît, d'une part, que la physique ne nous fait pas connaître l'ultime réalité indépendante mais « doit se contenter de décrire les phénomènes ». Cet auteur établit ce point par ses analyses rigoureuses des différentes tentatives pour essayer de faire de la théorie quantique une théorie à objectivité forte (chapitres 10, 12 et 13), montrant en particulier que les tentatives pour interpréter le processus de mesure comme un processus physique font toujours appel, souvent de façon cachée, à un principe de limitation de l'information que peut acquérir l'observateur, ou encore en évoquant la cohérence de l'interprétation épistémique de l'opération de mesure proposée par Bohr (chap. 11, par exemple). Cependant, cet auteur a néanmoins soutenu la position réaliste suivante qui se fonde sur l'idée que *la réalité indépendante devrait être dotée de certaines propriétés caractéristiques du domaine quantique, en particulier qu'elle serait non-locale et non atomisable* (c'est à dire non « analysable en une multitude d'éléments premiers »[50]). Concentrons-nous ici sur la non-localité (ou l'« enchevêtrement ») de la réalité indépendante qui joue aussi un rôle essentiel dans les propositions mentionnées ci-après. Pour d'Espagnat, ce serait la propriété d'intrication des états quantiques qui permettrait de justifier que la réalité indépendante est « un tout fortement enchevêtré » qui ne serait pas immergé dans l'espace-temps:

> … si l'on postule que la réalité indépendante est correctement et exhaustivement décrite par la physique quantique (…), on doit concevoir cette réalité indépendante comme un tout *fortement enchevêtré* ;[51] avec pour conséquence qu'il est impossible de concevoir des parties de ce tout occupant chacune un lieu défini.

Par ailleurs, comme nous l'avions mentionné ci-dessus (Question (a)), Michael Esfeld,[52] soutient une interprétation réaliste (non naïve) de la fonction d'onde dans le sens où cette dernière décrirait le « fait physique concret » que sont les *relations* entre les éléments de l'ontologie primitive, à savoir les corpuscules classiques réelles de la mécanique Bohmienne. Par conséquent, la non-localité est conçue comme *une propriété de la réalité indépendante* qui serait encodée dans la

---

[49] Bernard d'Espagnat, *Le Réel Voilé*, 134.

[50] D'Espagnat, *Le Réel Voilé*, 336. Cette propriété de la réalité indépendante serait, selon d'Espagnat, justifiée par le fait que dans la théorie quantique des champs, les phénomènes de création et d'annihilation de particules sont représentés par le changement d'état d'un *unique* vecteur dans l'espace des états approprié (un espace de Fock).

[51] D'Espagnat, *Le Réel Voilé*, 334. C'est nous qui soulignons le point essentiel de cette affirmation.

[52] Esfeld, "How to Account for Quantum Non-locality."

fonction d'onde. Enfin, dans le même ordre d'idées, Dorato[53] propose de considérer l'intrication comme une caractéristique intrinsèque de la réalité indépendante puisqu'il soutient l'idée selon laquelle cette propriété définirait le « nouvel ordre fondamental de la *nature*».[54] Un engagement ontologique qui paraît, en fait, outrepasser le changement radical de stratégie *explicative* de la physique qu'il constate :

> … rather than trying to *explain*[55] the quantum non-local correlations in terms of the causal ontology of the previous, classical theory of the world of our experience, we ought to accept them as paradigmatic examples of a new fundamental order of nature.

Si l'observation de corrélations non-locales, typiques du nouveau paradigme quantique, peut servir de fondement à l'explication des phénomènes physiques, pourquoi la non-localité serait-elle inscrite dans la « nature », comme propriété intrinsèque de la réalité indépendante ? Pourquoi la recherche de cohérence et d'efficacité *explicative* mènerait-elle à l'engagement ontologique ?

Pour ces trois auteurs, il serait ainsi possible d'assigner à la réalité indépendante les particularités des phénomènes quantiques et, en particulier, celle de non-localité qui trouve son expression formelle dans l'intrication des états. Cependant, si nous nous référons à la définition stricte de ces propriétés qu'en donne la théorie quantique standard (voir ci-dessus) il est facile de voir que ces dernières sont fortement dépendantes des observables choisies pour décrire le système considéré et qu'elles doivent donc être considérées comme des caractéristiques *épistémiques* du processus d'observation et non comme celles d'une hypothétique réalité-en-soi. Nous nous focaliserons ici sur la propriété d'intrication qui peut être considérée comme le trait le plus typique du domaine quantique auquel peuvent être reliées d'autres propriétés-clés, comme c'est le cas pour la complémentarité.[56]

Lorsque nous disons que l'état d'un système est « intriqué », nous supposons en fait deux choses. Tout d'abord que cet état est représenté par un vecteur d'un *sous-espace* de l'espace de *tous* les états possibles associé au système considéré qui

[53] Mauro Dorato, "Laws of Nature and the Reality of the Wave Function,"*Synthese* 192 (2015): 3179–3201.

[54] Dorato, Conference on "Causal Explanations, Structural Explanations and Entanglement," held at IHPST (Paris), 4 April, Metascience Seminar.

[55] Dorato, "Laws of Nature," 3179. Nous soulignons aussi.

[56] Par exemple, le théorème de Landau ("Experimental Tests of General Quantum Theories,"*Letters in Mathematical Physics*, 1987) permet de relier la propriété d'intrication de l'état d'un système composé de deux sous-systèmes avec la complémentarité de leurs observables respectives.

est engendré par les vecteurs propres *d'un jeu donné d'observables compatibles.* Par exemple, lorsque nous affirmons que le système composé S1 + S2 se trouve dans l'état intriqué :

$$| W > = 1/\sqrt{2}(| 0_1\, 0_2> + | 1_1\, 1_2>),$$

nous ne nous intéressons qu'aux valeurs d'une observable jointe, que nous noterons A1 ⊗ A2, où A1 et A2 sont deux observables dichotomiques à valeurs (0 ou 1) non dégénérées définies respectivement sur S1 et S2 et de vecteurs propres $|0_1>$, $| 1_1>$ (pour A1) et $| 0_2>$, $| 1_2>$ (pour A2). Par exemple, S1 et S2 peuvent être deux particules alors qu'A1 et A2 peuvent être des observables de spin selon deux directions spatiales (différentes ou non). Cependant, ce même système S (de deux particules, ici) *a d'autres propriétés* auxquelles nous ne nous intéressons pas pour spécifier son état, comme par exemple leurs positions, leurs impulsions ou leurs « couleurs » s'il s'agit de quarks, et prendre ces propriétés en considération mène à caractériser l'état de S *par un autre vecteur* de l'espace-produit des sous-espaces engendrés, respectivement, *par toutes les propriétés considérées de S1 et S2*. Ce dernier état peut très bien ne pas être intriqué relativement à la position, par exemple, si les positions des particules S1 et S2 ne sont pas corrélées (et même si leurs spins le sont). Il suffit donc de prendre en compte d'autres observables compatibles avec celles considérées initialement et dont les distributions de probabilités ne sont pas corrélées pour démentir l'idée que l'intrication est une propriété absolue de l'état d'un système. La propriété d'intrication d'un état, et donc les phénomènes de non-localité qui lui sont associés, sont toujours *relatifs au jeu d'observables choisi pour le décrire*, *ils ne sont pas absolus et ne peuvent donc pas être attribués à la réalité indépendante.*

En outre, même si nous ne nous intéressons qu'aux descriptions d'un système utilisant *à la fois toutes les observables compatibles* pouvant être définies sur ce système, la propriété d'intrication de son état resterait encore relative à la décomposition de son espace de représentation, c'est à dire à la façon dont est utilisé ce jeu d'observables. En effet, il est possible de montrer que l'état d'un système peut être intriqué relativement à une certaine décomposition de son espace de représentation (et donc au jeu d'observables permettant de spécifier ces sous-espaces), mais séparable relativement à une décomposition différente de cet espace, c'est à dire à un jeu différent d'observables définies sur ce même système. Donnons un exemple simple de cette relativité de l'intrication tiré du livre de Rieffel et Polak[57] sur l'informatique quantique en insistant ici sur le lien entre la

[57] Eleanor Rieffel and Wolfgang Polak, *Quantum Computing: A Gentle Introduction* (Cambridge, Massachusetts: MIT Press, 2011).

décomposition de l'espace de représentation et le choix du jeu d'observables considérées.

Soit un système S composé de quatre particules et soient A1, A2, A3 et A4 quatre observables dichotomiques (dont les valeurs ne sont pas dégénérées) définies, respectivement, sur ces quatre particules. Il peut s'agir, par exemple, d'observables de spin selon une direction spatiale. En utilisant les mêmes notations que ci-dessus, supposons que l'état de ce système puisse s'écrire comme :

$$|W> = 1/2(|0_1\,0_2\,0_3\,0_4> + |0_1\,1_2\,0_3\,1_4> + |1_1\,0_2\,1_3\,0_4> + |1_1\,1_2\,1_3\,1_4>).$$

Cet état peut être dit « intriqué » *si* l'espace de représentation H est décomposé comme[58] :

$$H = H1 \otimes H2 \otimes H3 \otimes H4,$$

où chacun de ces sous-espace est engendré par les deux vecteurs propres des quatre observables A1, A2, A3 et A4 – c'est à dire que les vecteurs de ces sous-espaces peuvent s'écrire comme des 1-qubits ($a_i|0_i> + b_i|1_i>$), avec i prenant les valeurs 1, 2, 3 et 4. En effet, on peut montrer que dans ce cas *il n'existe pas* de nombres a_i et b_i tels que $|W>$ soit le produit de quatre vecteurs appartenant, respectivement, à chacun de ces sous-espaces.

Cependant, si nous considérons maintenant *les observables jointes A1 ⊗ A3 et A2 ⊗ A4,* respectivement définies sur les sous-espaces H1 ⊗ H3 et H2 ⊗ H4, *ce même état* $|W>$ *peut être factorisé* de la façon suivante :

$$|W> = |W_{1,3}> \otimes |W_{2,4}>,$$

où $|W_{1,3}> = 1/\sqrt{2}(|0_1\,0_3> + |1_1\,1_3>)$ est un vecteur de H1 ⊗ H3, relatif à l'observable jointe A1 ⊗ A3, et $|W_{1,3}> = 1/\sqrt{2}(|0_2\,0_4> + |1_2\,1_4>)$ est un vecteur de H2 ⊗ H4, relatif à l'observable jointe A2 ⊗ A4.

Ces deux arguments montrent ainsi que la propriété d'intrication de l'état d'un système, et donc les phénomènes de non-localité qui lui sont associés, *sont relatifs au jeu d'observables qui a été choisi pour le décrire*. Par conséquent, les propriétés typiquement quantiques ne peuvent, pas plus que les propriétés classiques ou relativistes, être considérées comme absolues, indépendantes de nos moyens d'investigation, et ne peuvent donc être assignées à une hypothétique réalité indépendante.

[58]En fait, comme le remarquent Rieffel et Polak, dire qu'un état de ce type (caractérisé par des variables dichotomiques) est intriqué sans préciser par rapport à quel choix d'observables se rapporte cette propriété n'est en général qu'un raccourci (trompeur) pour dire que c'est par rapport à la décomposition de l'espace de représentation en sous-espaces bidimensionnels (dont les vecteurs sont des 1-qubits).

4. Conclusion

En plus des arguments généraux évoqués à la section 2 à l'encontre du réalisme scientifique, les arguments spécifiques présentés ci-dessus (section 3) montrent que le projet d'une métaphysique fondée sur les spécificités du domaine quantique n'est pas tenable. En effet, dans le prolongement de l'argument « positif » général relatif à la construction de l'objectivité que nous avons présenté dans la section 2, ces arguments spécifiques rendent particulièrement explicite, grâce à la notion centrale d' « observable », l'idée que les phénomènes du domaine quantique sont « mis en forme » par des procédures théorico-expérimentales qui sont clairement dépendantes de nos moyens d'investigation et de nos représentations. Par conséquent, l'idée qu'ils pourraient nous révéler les structures intrinsèques de la réalité indépendante relève de la croyance, sans doute stimulante pour certains comme visée régulatrice, mais ne repose, selon nous, sur aucune justification crédible.

REAL KNOWLEDGE UNDERMINING LUCK

Raphael VAN RIEL

ABSTRACT: Based on the discussion of a novel version of the Barn County scenario, the paper argues for a new explication of knowledge undermining luck. In passing, an as yet undetected form of benign luck is identified.

KEYWORDS: epistemic luck, environmental luck, justification, method of belief-formation

1. Introduction

It is widely assumed that knowledge is incompatible with several types of epistemic luck. In Gettier cases,[1] a subject is *lucky* to arrive at a true belief when inferring a truth from justified but false beliefs and, *therefore*, does not acquire knowledge. In Russell's scenario of a "man who looks at a clock which is not going, though he thinks it is, and who happens to look at it the moment when it is right,"[2] the observer forms a true belief about time, but is lucky when doing so and, *because* he is lucky, does not acquire knowledge. Now, as is well known, there are various forms of benign luck, i.e. forms of luck that do *not* interfere with knowledge acquisition. Any account of knowledge-undermining luck will thus have to specify further conditions in order to capture the type of luck epistemologists are after.

The present paper argues for one particular explication of the sort of luck epistemologists typically regard to be incompatible with knowledge, and it discusses the relation between this explication and explications proposed by Duncan Pritchard, Masahira Yamada, and Mylan Engel. In a nutshell, I will argue that the sort of luck that seems to interfere with knowledge should be relativized not to the method of belief formation, as, for instance, Pritchard and Yamada would have it, but, rather, to *whatever* supports the person's belief – the method employed and the implicit or explicit beliefs that may support the belief.[3]

---

[1]Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121–123.

[2]Bertrand Russell, *Human Knowledge. Its Scope and Limits* (London: Routledge, 2005 (1948)), 170.

[3] Note that I do not intend to defend the view that based on this characterization, we may arrive at a definition of knowledge; there may be constraints on knowledge that cannot be cashed out in terms of an anti-luck condition.

© LOGOS & EPISTEME, VII, 3 (2016): 325-344

In Section 2, I will introduce and discuss a novel version of the Barn County-scenario, and suggest that given some additional assumptions about the subject in the scenario, it seems intuitive that a subject may acquire knowledge in the presence of environmental luck. In Section 3, I will offer an argument that blocks one possible criticism of this result. In Section 4, I suggest that the novel version of the Barn County Scenario enables us to identify an as yet undetected form of benign luck. In Sections 5-8, I will, based on the novel Barn County scenario, develop an explication of the type of luck that seems to interfere with knowledge and discuss its relation to anti-luck conditions offered by Engel and Pritchard. Whereas Pritchard's condition is in need of a refinement, one may regard the argument developed below to support Engel's characterization, according to which luck should be relativized to the evidence a subject has for a belief – depending on the intended interpretation of 'evidence.'

2. Two Versions of [Barn County], a Difference, and an Intuition

This section introduces the classical and a novel version of the Barn County-scenario, suggests that based on intuitions about the novel version, environmental luck turns out to be compatible with knowledge acquisition, and comments on the difference between the two versions, suggesting that the difference explains the difference in knowledge-acquisition. Here is the classical version:

> [Barn County]
>
> Simon sees a barn in front of him. Simon forms the true belief that the thing in front of him is a barn. The causal chain leading from the fact that there is a barn to Simon's belief formation is perfect, unlike the environment. The barn he sees is the only real barn in an area where all other barn-like objects (and there are many) are mere barn facades, all indistinguishable, from Simon's perspective, from real barns. Simon was lucky. In this environment, Barn County, he might have been easily misled.[4]

It is a widely shared intuition that in scenarios of this sort, environmental luck interferes with knowledge – in this scenario, Simon does not know that there is a barn in front of him. Environmental luck will here be understood in a general sense as follows: things might very easily have not worked out for our subject (the subject is lucky) due to circumstances in the environment (which makes this type of luck a form of *environmental* luck); it is, to use Pritchard's terminology: not of

[4] Cf. Alvin Goldman, "Discrimination and Perceptual Knowledge," *The Journal of Philosophy*, 73 (1971): 771-791.

the 'intervening' sort.[5] In this sense, Simon is lucky. Moreover, Simon does not acquire knowledge, which is, at least in part, due to the fact that it was a matter of luck (in the general sense) that he arrived at a true proposition.

I will take this as a datum. But what if Simon had some non-decisive information that plays a justificatory role for the belief that there is a barn in front of him? Consider the following scenario:

> [Barn County*]
>
> Before traveling to Barn County, Simon* talked to his partner Martha who traveled to Barn County before. Martha, who spotted and examined the only real barn in Barn County, tells Simon* that there is a barn at a particular crossing or that there is a church at this particular crossing, and based on this disjunctive information, Simon* is justified to believe that there is a barn at this crossing or that there is a church at this crossing. Note that Martha does *not* tell Simon* anything else; in particular, Simon* does not learn anything about the fact that there are many, many barn façades in Barn County. Simon* travels through Barn County. The first barn Simon* spots is the *only* real barn in Barn County (information Simon* does not possess); and it is a barn at the crossing indicated by Martha. Simon* does not have any reason to believe that there is also a church at this particular crossing.

Let us assume that the scenario is otherwise indistinguishable from the original [Barn County] scenario (so, for instance, Simon and Simon* employ the same method of belief formation etc.) I take for granted that, before traveling to Barn County, Simon* at best knows the disjunction that there is either a church or a barn at this particular crossing, but is not thereby in a position to know that there is a barn at this particular crossing. Do things change when Simon* spots the barn? Is Simon* in a position to know that there is a barn in front of him, using visual information alone when forming the belief?

It appears that Simon* is in a position to acquire knowledge when forming the belief based on visual information alone. Assume, first, that Simon* *uses* disjunctive information provided by Martha when forming the belief. Then, I think, he will clearly acquire knowledge (provided, of course, that he is able to distinguish a barn from a church). More importantly, however, it appears that he will acquire knowledge even when the information provided by Martha merely plays a justificatory role and does not enter the process of belief-formation. It seems that Simon* does not need to actualize the belief that there is a church or a barn at this particular crossing, or that Martha told him so. When pressed, Simon*

[5]Duncan Pritchard, "Knowing the Answer, Understanding, and Epistemic Value," *Grazer Philosophische Studien* 77 (2008): 330.

might have asserted that there is a church or a barn at this particular crossing (and the fact that Martha told him so could have been causally relevant for Simon*'s assertion). Or, upon reflection, Simon* may have come to actualize this belief. Whether or not he does seems to be irrelevant for the question of whether or not he acquires knowledge in [Barn County*]. Simon* acquires knowledge in [Barn County*] because he *possesses* additional information, not because he actually *uses* it when forming the belief.

At the same time, it appears that Simon* was *lucky*, in a way similar to the way Simon is lucky in [Barn County]. Relying on a modal interpretation of luck: Simon was lucky because, had he spotted a barn façade, he would have formed a false belief, or in *most* nearby possible worlds, he would have formed a false belief. In this respect, Simon* was just as lucky as Simon.

Here, I merely report my intuitions. As the paper proceeds, I will present two indirect arguments for the claim that Simon* acquires knowledge in [Barn County*]. First, however, let me comment on the type of information Simon* possesses in [Barn County*] and, hence, on the difference between the two versions of the scenario.

For the sake of simplicity, I will refer to the information Simon* possesses in [Barn County*] as 'background information,' without presupposing any technical, or theory-laden notion of a *background*. To get a better idea of background information, let us briefly reflect on some structural features of the information provided by Martha in [Barn County*].

One may tentatively describe background information in terms of what it is *about*. Background information, in [Barn County*], concerns the content of the belief Simon* forms in [Barn County*], or a proposition relevantly related to this content, namely, that there is a barn at this particular crossing. In contrast, in [Barn County], Simon does not possess any particular information about the target proposition, although, of course, he will have to possess information about barns, or the concept of a barn, in general (or so it is tacitly understood). In addition, the information Simon* possesses in [Barn County*] should not itself be generated under conditions of environmental luck. Otherwise, it is not so clear whether information provided by Martha really does the trick in [Barn County*]. And, as I have already stressed, we should conceive of Simon*'s background information so that it does not play any role in the formation of the belief that there is a barn in front of him. It is not that Simon* fails to make the connection; it is just that he bases his belief on visual information alone. You may, but need not, think of background information as *non-occurrent, implicit,* or *non-actualized.* In [Barn County*], Simon* need not actualize the belief that there is a church or a barn at

this particular crossing, or that Martha told him so (in order to form the belief). Since Simon* does not use background information in the process of belief-formation, the belief may have been implicit, non-occurrent, or non-actualized. Then, if background information plays a role for knowledge-acquisition, Simon* acquires knowledge in [Barn County*] because he *possesses* additional information, not because he actually *bases his belief on* it, or *uses* it when forming the belief. In this sense, background information is *background* information (as I use the term here). These features, together with the example just presented, should provide a sufficiently clear understanding of the type of information whose presence distinguishes [Barn County*] from [Barn County].

Based on these tentative characterizations, we are now in a position to mount an argument for the view that background information *can* make a difference to the question of whether or not a subject acquires knowledge, thereby offering a first indirect argument for the claim that in [Barn County*], Simon* acquires knowledge – namely, by blocking a possible counter-argument.

3. An Indirect Argument: Background Information and Justification

Background information can play a role for knowledge acquisition. Consider two persons, Sarah and John, who, together, overhear a conversation among two people they do not have any additional information about; in particular, they do not have any reason to believe that the short exchange they overhear is sincere, or aims at truth, nor do they possess evidence to the contrary. They merely hear one sentence: "When you cross a horse with a zebra, chances are dim that the offspring will be able to reproduce." They both form the belief that this is so, based in their overhearing this snippet of a conversation. Sarah, unlike John, possesses background information that may play a justificatory role for the belief that when you cross a horse with a zebra, chances are dim that the offspring will be able to reproduce. For instance, we may assume that she knows that when you cross a horse with a donkey, chances are dim that the offspring will be able to reproduce, and that the relation between donkeys and horses is similar to the relation between horses and zebras. John does not possess this information, nor does he possess any similar information. Then, I would say, whereas Sarah may have acquired knowledge in this scenario, things are less clear with John. There are numerous differences between the two: Whereas Sarah is in a position to integrate the information into her body of belief and is in a position to make the connection, John isn't. Moreover, John's belief would not be as stable as Sarah's belief. On the assumption that Sarah and John are equally rational, they may react differently when presented with the information that the person who uttered the

sentence is a notorious liar; Sarah will not be irrational when she sticks to her belief. Not so John – he should abandon his belief when being told that his source was a notorious liar.

Of course, this does not show that Simon* acquired knowledge in [Barn County*]; but it offers a response to the worry one might have that implicit background information cannot *possibly* make a difference in the context of knowledge acquisition. Prima facie, background information can play such a role. The argument shows that the intuition that Simon* acquired knowledge in [Barn County*] is, if misguided, not misguided because it credits background information with a role it cannot possibly play.

If you take knowledge or justification to be tied to cognitive achievements, or virtues, or to any form of *process* of belief formation, you may feel reluctant to accept the result that Simon* acquires knowledge in [Barn County*], or, for that matter, that Sarah acquires knowledge in the situation just sketched. How can information Simon* does not *use* when forming the belief bear on the question of whether or not the belief amounts to knowledge? The relation between the conclusion that Simon* acquired knowledge in [Barn County*] and these views does not seem to be straightforward. First, there *is* a process of belief formation that may fit some of the bills (for instance, be reliable). Background information is an extra. And at least, it is not obvious that, by subscribing to some form of reliabilism, or virtue epistemology, one is committed to the claim that beliefs that are not used in belief-formation cannot play any *additional* justificatory role. We will turn back to this point in Section 7.

So, it appears that one argument one might want to raise against the intuition that Simon* acquires knowledge in the scenario fails. Before turning to the discussion of how this result bears on knowledge-undermining luck, let me briefly discuss an interesting feature of the scenario – [Barn County*] involves a novel form of benign luck.

4. Benign Luck in [Barn County*]

There are innocent, or benign forms of luck, forms of luck that are assumed to be compatible with knowledge acquisition. Following Pritchard's[6] interpretation of Unger's reflection on varieties of epistemic luck,[7] one can distinguish three forms of benign luck:

(1)It is a matter of luck that the proposition known is true.

[6]Duncan Pritchard, "Epistemic Luck," *Journal of Philosophical Research* 29 (2004): 191-220.
[7]Peter Unger, "An Analysis of Factual Knowledge," *Journal of Philosophy* 65 (1968): 157-170.

(2)It is a matter of luck that the agent is capable of knowledge.

(3)It is a matter of luck that the agent acquired the evidence that supports her knowledge.[8]

An example for the first type of luck is this: You witness a car accident. That it is true that there was a car accident is lucky, in the sense that things could easily have been different (this is supposed to follow from the idea of an accident). As for (2), consider a scenario where a subject acquires knowledge, but could have easily ceased to exist due to circumstances present in the environment. In both cases, or so it seems, luck does not interfere with knowledge. Finally, one may be extremely lucky that one gathered the evidence one has for a belief. When a bank teller sees the robber slip the mask for a short moment and recognizes the robber, gathering of evidence may very well count as lucky.[9]Again, it appears that once one has acquired evidence, one is in a position to acquire knowledge, independent of whether evidence acquisition was a matter of luck.

The taxonomy of types of epistemic luck in (1)-(3) distinguishes types of luck in terms of the *object* or *target* of luck; the truth of the proposition (1), the ability to acquire knowledge (2), or the availability of evidence (3). Now, it is clearly a matter of luck that *the piece of information provided by Martha became relevant in [Barn County*]*, that it did play a justificatory role in the scenario. Simon* might, very easily, have looked at a barn façade. If he had looked at a barn façade, information provided by Martha would not have played any justificatory role at all.

This form of luck does not collapse into any of (1)-(3); Simon* is lucky in [Barn County*], but not because it is a matter of luck that the proposition that there is a barn in front of him is true, or because it is a matter of luck that he is capable of knowledge or belief-formation, or because it is a matter of luck that he acquired background information or visual information. We can add the following type of luck to our taxonomy of benign forms of epistemic luck:

4. It is a matter of luck that part of the information a subject possesses plays a justificatory role for the belief that *p*.

This form of luck resembles the lucky occurrence of evidence, (3), although here, it is not the *acquisition* of evidence that is lucky, but, rather, the fact that in a context, information the subject already possessed *becomes* evidence, or played a

[8] For a discussion of the relation of this condition to doxastic luck, cf. Pritchard, "Epistemic Luck"; in the present context, the relation between the two does not matter.

[9] Cf. Robert Nozick, *Philosophical Explanations* (Cambridge (MA): Harvard University Press, 1981).

justificatory role. Note, however, that the fact that there is an as yet undetected form of benign luck involved in [Barn County*] surely does not explain why Simon* acquired knowledge. So, let us turn back to the main topic of the paper: Which form of luck is incompatible with knowledge?

5. Relativizing Luck, the Method of Belief-formation, and Two Desiderata

The type of luck that is usually regarded as problematic with respect to knowledge acquisition concerns the fact that a subject ended up with a true belief. Call this form of luck 'resultant luck.'[10] We have seen that [Barn County*] does involve resultant luck: In the relevant sort of environment, Simon* was lucky that he acquired a true belief. By these lights, resultant luck appears to be compatible with knowledge acquisition. However, resultant luck needs to be *relativized* in order to yield the sort of luck that is incompatible with knowledge (as has been argued, for instance, by Engel[11] and Baumann.[12]) This section argues that resultant luck, when relativized to the method of belief formation does not, *pace* Pritchard,[13] contradict knowledge acquisition. Thus, resultant luck with respect to the method of belief-formation is not incompatible with knowledge.

According to Pritchard, knowledge requires that the acquisition of a true belief was not lucky with respect to the *method* employed when forming the belief.[14] Transformed into an anti-luck condition on knowledge, and ignoring, for the moment, Pritchard's particular interpretation of luck in modal terms, this reads as follows:

> [Condition$_{\text{Method}}$]
>
> *x* knows that *p* only if it is not just a matter of luck, given *the method* of *x*'s belief formation, that *x*'s belief that *p* is true.

Given Simon*'s way of belief formation, Simon* *was* just as lucky to arrive at a true belief in [Barn County*] as Simon was in [Barn County]. Couched in modal terms: For both scenarios, it seems that in most nearby possible worlds where Simon*/Simon bases his belief on visual information, he ends up with a false belief. Same method, yet in [Barn County*], Simon* ends up with knowledge. In the light

[10]Peter Baumann, "No Luck with Knowledge? On a Dogma of Epistemology," *Philosophy and Phenomenological Research* LXXXIX (2014): 525.
[11]Mylan Engel, "Is Epistemic Luck Compatible with Knowledge?" *The Southern Journal of Philosophy* 30 (1992): 59-75.
[12]Baumann, "No Luck with Knowledge?".
[13]Duncan Pritchard, *Epistemic Luck* (Oxford: Oxford University Press, 2005).
[14]Pritchard, *Epistemic Luck*.

of [Barn County*], Pritchard's explication appears to be mistaken.The same appears to hold for related explications. Yamada offers a more detailed explication of luck that is relativized to the method of belief-formation (which he then goes on defending as a necessary condition on knowledge). When taken as an explication of knowledge-interfering epistemic luck, his first characterization (later in his paper fleshed out in more detail) suggests that belief-acquisition is not relevantly lucky[15] if and only if

1. the method M used is truth-conducive
2. it is not an accident that one correctly applied M
3. it is not an accident that one is using a truth-conducive method.[16]

Yamada suggests that his account delivers the correct result for [Barn County] - the environment in Barn County ensures that it is an accident that the subject correctly applies the method of belief-formation. Yamada offers an interesting account of the method the subject applies in this context that delivers the correct result: On this account, the method the subject employs is too easy to misapply. Hence, condition 2 is not met. In [Barn County *], the subject, by assumption, applies the same method. Still, it appears that the subject is in a position to acquire knowledge. [Barn County*] constitutes a counter-example to method-relativized accounts of luck.

But maybe, this was too quick. One may want to object that background information provided by Martha has an impact on the method Simon* employs when forming his belief. However, by assumption, Simon* does not *use* background information when forming the belief that there is a barn in front of him. Background information may, here, be only implicit, or non-occurring. And given the following two conditions on the notion of a method of belief formation that seem to characterize the notion of a method Pritchard has in mind, background information does not have an impact on the method of belief formation in [Barn County*] either. First, Pritchard characterizes the method as a 'way of' forming a belief.[17] The ways Simon and Simon* form their beliefs in [Barn County] and [Barn County*] respectively, are the same – they look at a barn and,

[15] In fact, Yamada suggests that when these conditions are met, there is "no sense in which it is an accident that [the subject] correctly believes [whatever it believes]" (Yamada, "Getting It Right by Accident," 82). If cases of benign luck discussed above make for correct though accidental belief, this is in need of further elaboration.

[16]Masahiro Yamada, "Getting It Right by Accident," *Philosophy and Phenomenological Research* LXXXIII (2011): 82.

[17]Pritchard, *Epistemic Luck*, 163.

based on visual information, form the belief. Thus, by these lights, background information has no impact on the method of belief-formation. Second, little reflection on the cases that have inspired reference to the method of belief-formation reveals that the method of belief formation is supposed to concern the means by which the belief is formed, so that again, background information, as present in [Barn County*], does not have any impact on the method of belief formation. The condition is supposed to rule out cases like Russell's clock and Barn County.[18] By these lights, it turns out that reference to a method of belief formation is not designed to cover background information.

As a consequence, we should reject Pritchard's claim that this type of luck is incompatible with knowledge acquisition. But, obviously, luck with respect to the method of belief formation *may* interfere with knowledge, as, for instance, [Barn County] seems to indicate. We are thus faced with two desiderata any successful account of knowledge-undermining resultant luck has to meet:

[D-1] An account of knowledge-undermining luck should explain the difference between [Barn County] and [Barn County*].

[D-2] An account of knowledge-undermining luck should explain why sometimes, though not always, luck with respect to the method of belief formation *does* interfere with knowledge.

6. Meeting the Desiderata: an Anti-luck Condition

Let us proceed in a piecemeal fashion. Note, firstly, that there is one relativization of resultant luck that yields the desired result, but lacks a number of other theoretical virtues. In order to introduce this relativization, let us capture the thought that Simon, in [Barn County], does *not* possess independent information on the barn he spots, information that would be analogous to the information provided by Martha in [Barn County*], by saying that his background information is *empty*.

Then, it appears that *with respect to the method of belief formation and background information about the target proposition*, it was *not* a matter of luck that Simon* arrived at a true belief in [Barn County*]. At the same time, it *was* just a matter of luck, given Simon's (empty) background information *and* method of belief formation, that he ended up with a true belief in [Barn County]. This kind of relativization – relativization with respect to background information *and* method of belief formation – appears to track an interesting connection, as the following, prima facie plausible *explanations* seem to indicate:

[18]Pritchard, "Epistemic Luck," 207f.

1)In [Barn County*], Simon* acquires knowledge *because* with respect to his background information and the method he employs when forming the belief, it is not just a matter of luck that his belief is true.

2)In [Barn County], Simon does *not* acquire knowledge *because* with respect to his (empty) background information and the method he employs when forming the belief in that context, it *is* a matter of luck that his belief is true.

We thus arrive at an explanation of the difference between the cases. The intuition that there is a difference between the cases as regards to knowledge acquisition does not come out of the blue. If you find these explanations compelling, but are not entirely sure about the intuition that in [Barn County*], Simon* acquired knowledge, you may regard the plausibility of these explanations as constituting another indirect argument for the claim that in fact, Simon* has acquired knowledge in this scenario: Given some relevant aspects of his cognitive system, it was not purely a matter of luck that he arrived at a true belief in [Barn County*]; and with respect to the same aspects, he was lucky in [Barn County]. The intuition that he acquired knowledge may, thus, in fact track an important distinction, that makes for a relevant difference between the two cases.

Based on (1) and (2), we also come to see why relativization to the method of belief-formation alone will sometimes, but not always, pick out the right form of luck: Relativization to the method of belief-formation *and* background information is equivalent to relativization to the method of belief formation alone, if the set of background information of the subject is empty. We thus meet the two desiderata.

This form of relativization, however, falls short of a general account of knowledge-undermining luck, for two reasons. First, we lack a general idea of what may constitute background information, and, second, just lumping background information and method of belief formation together in order to arrive at a disjunctive relativization may seem *ad hoc*. Although we may be on the right track (we have an account that meets our desiderata for the particular cases), we still lack an account that meets further conditions of theoretical elegance.

Now, if we were able to come up with an explanation of what ties background information and the method of belief-formation together, so that they, together, turn out to form the relevant parameters in question, we might be able to offer a general and non-disjunctive characterization of the relevant type of epistemic luck. Fortunately, it seems that there is a straightforward way of doing so. Both, background information and method of belief formation play a justificatory role with respect to the subject's belief. Why not lump these two

together, and explicate the relevant form of luck in terms of a relativization to the subject's *justification base* for the target proposition?

7. Knowledge Undermining Luck

The basis for justification for a proposition may include background information that does not, on any plausible reading, enter the method of belief formation. At the same time, the justification base may include the method of belief formation. Since it is not entirely clear to me whether one would thereby depart from a standard interpretation of *evidence*, I prefer speaking of a *justification base*. Some will assume that it involves evidence only, others might allow reliable processes to be part of the justification base. The neutrality of 'justification base' is, in the present context, an advantage. And it appears to be a notion that is clear enough: A subject's justification base for a belief that *p* is the sum total of what bears on the belief that *p* (including, of course, counter-evidence). Further explications can be deferred to theories of evidence or justification. We arrive at the following anti-luck condition on knowledge:

> [Condition$_{\text{Justification-Base}}$]
>
> *x* knows that *p* only if, with respect to the *justification base* for *p*, it was not just a matter of luck that *x*'s belief that *p* was true.

The so refined condition appears to capture the idea that an anti-luck condition amounts to well-foundedness of the belief, not only in terms of method, but *also* in terms of what may count as belonging to the justification base, i.e. in terms of well-foundedness in the "conjunction" of method and background. In [Barn County*], Simon* is not lucky with respect to background-information *together with* method of belief-formation, though he is lucky with respect to the method of belief-formation alone. The condition meets the desiderata: With respect to his justification base (method and background information) in [Barn County*], it is not just a matter of luck that Simon* arrived at a true belief, with respect to his justification base in [Barn County], it is a matter of luck that Simon arrived at a true belief, and this is so because there is a difference in the justification base in the two cases. Moreover, the account is *general*, and it offers a *unified* account of what ties the method of belief formation and background information, or information relevantly similar to the information provided by Martha, together.

Now, compare this condition to the anti-luck condition one can arrive at when departing from Engel's characterization of veritic luck:

(VL) A person *S* is *veritically lucky* in believing that *p* in circumstances *C* if and only if, given *S*'s evidence for *p*, it is just a matter of luck that *S*'s belief that *p* is true in *C*.[19]

As has been pointed out by Pritchard and Smith, (VL) clearly does not amount to an explication of *being lucky*.[20] Nevertheless, it might very well be a principle that governs the sort of luck that is incompatible with knowledge, as follows:

[Condition$_{\text{Evidence}}$]

x knows that *p* only if it is not just a matter of luck, given *x*'s *evidence* for the belief that *p*, that *x*'s belief that *p* is true.

Obviously, if the justification base for a belief coincides with the evidence for this belief, [Condition$_{\text{Evidence}}$] and [Condition$_{\text{Justification-base}}$] are equivalent. Then, what has been said so far would turn out to be an argument for an explication of an anti-luck condition in terms of veritic luck, as defined by Engel. Maybe, this is what Engel had (and has) in mind – he does not comment much on the notion of evidence he presupposes. If, however, the method of belief-formation is not to be included in the evidence a person has, then [Condition$_{\text{Evidence}}$] differs from [Condition$_{\text{Justification-Base}}$], and the latter will offer the correct result in cases where, say, a person has some evidence for a target proposition, is not lucky with respect to the evidence when forming the belief, but is lucky with respect to the method of belief-formation in a way that interferes with knowledge, independent of the additional evidence the person has. When a subject arrives at a true belief by wishful thinking, ignoring all the positive evidence she has for that belief or target proposition, she does not acquire knowledge. Given her justification-base, including method *and* evidence, it was a matter of luck that she arrived at true belief. Given her evidence alone (on a reading that does not include the method of belief-formation) it was not.

Another advantage of [Condition$_{\text{Justification-Base}}$] is that it is independent of our particular views regarding the kind of justification required for knowledge. The reliabilist and the evidentialist alike may accept that the reliability of the process of belief formation *and* the evidence available to a subject may play a justificatory role for a given belief. And they may accept that a subject can be lucky when

[19]Mylan Engel, "Epistemic Luck," in *A Companion to Epistemology* (2nd edition), ed. Jonathan Dancy, Ernest Sosa and Matthias Steup, 336-339. London: Blackwell, 337; similarly in Engel, "Is Epistemic Luck Compatible with Knowledge?", 67.

[20]Duncan Pritchard and Matthew Smith, "The Psychology and Philosophy of Luck," *New Ideas in Psychology* 22 (2004): 1-28.

acquiring a true belief with respect to (i) the available evidence, (ii) the process of belief formation, and (iii) with respect to the available evidence together with the process of belief formation. Why shouldn't they agree that it is the conjunction of the two that offers the relevant parameter to relativize knowledge undermining luck? As such, reliabilism and evidentialism can remain neutral on this point. Reliablists and evidentialists disagree on how the notion of justification, *as required in a characterization of knowledge*, should be spelled out. And it is not obvious that considerations concerning this latter problem should directly bear on explications of knowledge-undermining luck. By the lights of [Barn County*], it seems that an explication of knowledge undermining luck may require a notion of justification that encompasses *both* types of belief-support – method of belief-formation and available evidence (including background information).

This is not a merely terminological point. The condition we use to identify the relevant relativization of knowledge undermining luck is *conceptually* independent of the various candidate definitions of the sort of justification required for knowledge. We can judge that in [Barn County*], Simon* is not lucky with respect to all the things that support his belief, *whatever* belongs to these – evidence, a process of belief formation etc. Hold these fixed, and it is not just a matter of luck that Simon* arrived at a true belief. It is not the job of a theory of knowledge-undermining luck to offer a full-blown theory of the type of justification allegedly required for knowledge.

Reflecting on the question of what it is to hold the justification base fixed, in the context of modal explications of an anti-luck condition on knowledge, will offer a more thorough understanding of what belongs, and what does not belong to the justification base for a belief.

8. Beliefs and Their Justification Base

Pritchard suggests a modal interpretation of luck. He describes the connection between luck and knowledge as follows:

> For all agents, ø, if an agent knows a contingent proposition ø, then, in nearly all (if not all) nearby possible worlds in which she forms her belief about ø in the same way as she forms her belief in the actual world, that agent only believes that ø when ø is true.[21]

Let us first try to arrive at a less baroque version of this explication; it appears to be unnecessarily complex. Unless one can form a belief without believing it, and I don't see how one could do that, Pritchard's explication is

[21] Pritchard, *Epistemic Luck*, 163.

equivalent to the following (if we restrict quantification to contingent propositions):

> For all agents, propositions, if the agent knows the proposition that *p* then in nearly all (if not all) nearby possible worlds in which she forms the belief that *p* in the same way as she forms her belief that *p* in the actual world, it is true that *p*.

According to Pritchard, knowledge will have to meet the following condition:

> [Condition$_{\text{Method}}$]
>
> *x* knows that *p* only if in nearly all (if not all) nearby possible worlds in which she forms the belief that *p* in the same way as she forms her belief that *p* in the actual world, it is true that *p*.

The upshot is that when we assess whether or not a subject was lucky in the relevant respect, we should not only check some arbitrary counterfactual scenarios; we should check those counterfactual scenarios where the subject forms her belief based on the same method. If you believe Pritchard's modal account of luck to be illuminating, you might consider the following condition to offer a further illumination of the condition proposed above:

> [Condition$_{\text{Justification-base MODAL}}$]
>
> *x* knows that *p* only if in nearly all (if not all) nearby possible worlds in which her justification base for the belief that *p* is the same as it is in the actual world, it is true that *p*.

If we take the method of belief formation to be always included in the justification base, we need not mention in addition that the subject believes that p; if there is a way of belief-formation, there is the resultant belief. I think that this condition clearly draws the line where it should do: Simon, in [Barn County], does not know that there is a barn in front of him because the support his justification base lends to his belief that there is a barn in front of him in a counterfactual scenario where he spots a barn façade, is the same as it is in the actual scenario. On the other hand, in [Barn County*], in all nearby possible worlds where Simon*'s justification base supports his belief in the same way as it does in the actual scenario, the content of his belief will be true. Since a justification base, as understood here, may cover the method of belief formation, we can be sure that all cases Pritchard wants to cover and where additional justification does not play any role at all, are covered by this condition as well.

Note that this imposes a condition on *belonging to the justification base for a proposition* according to which background information provided by Martha

does *not* belong to the justification base for Simon*'s belief that there is a barn in front of him in nearby worlds where Simon* looks at a barn façade. I think that this matches our intuitive judgments, and it also makes sense if we interpret the notion of a justification base, just to illustrate the point, in probabilistic terms; then, a proposition that *q* belongs to a subject's justification base *b* for the proposition that *p* only if it makes a difference, positive or negative, to the conditional probability that the belief is true given at least one subset of the justification base *b*. The information provided by Martha does not make a difference to Simon*'s belief that there is a barn in front of him in counterfactual scenarios, where there is *no* barn in front of him. This articulates the idea that we hold fixed everything that either speaks in favor of or against the truth of the target proposition that *p*, and is relevantly related to the subject so that it bears on how the subject is justified with respect to the belief that *p*.

One may hope to bypass the problem of individuating the justification base for a belief by just considering the subject's total set of beliefs, and offering the following explication:

> [Condition$_{\text{Justification-Base*}}$]
>
> *x* knows that *p* only if in nearly all (if not all) nearby possible worlds in which her total set of beliefs is the same as it is in the actual world and lends the same support to the belief that *p,* it is true that *p*.

We just take the totality of x's actual beliefs, or the totality of information x possesses, (possibly including the method of belief-formation and other relevant factors) and consider worlds where this totality relates, in terms of justification or support, to the belief that *p* in the same way as it does in the actual world.

Unfortunately, [Condition$_{\text{Justification-Base*}}$] is subject to straightforward counterexamples. Assume that Simon* was not only told by Martha that there is a barn or a church at some particular crossing, but also, by one malevolent friend, that there were a barn or a church at one other crossing, and by yet another malevolent friend, that there were a barn or a church at yet another crossing, and so on…

It seems that Simon* may still know that there is a barn at the crossing in this version of [Barn County];[22] but his total set of beliefs will lend the same evidence to the belief that there is a barn in front of him in counterfactual

[22] The fact that he was told so many falsehoods should not affect the safety of his belief that Martha told him the truth; thus, Martha should have provided information in a context that is different from the context of belief formation based on false information by malevolent friends.

scenarios where he looks at barn façades. So, we will have to stick to a relatively demanding notion of a justification base a subject has for a proposition.

I have been told that one might get the impression that [Condition$_{\text{Justification-Base MODAL}}$] is equivalent to Pritchard's [Condition$_{\text{Method}}$]. But this is not the case. Based on the latter, we should conclude that Simon* does *not* know that there is a barn in front of him. Why is that? There are a number of nearby possible worlds where Simon* employs the same method of belief formation, but looks at a barn façade. Hence, he does not acquire knowledge. [Condition$_{\text{Justification-Base MODAL}}$] yields different results. In worlds where Simon* looks at a barn façade, his justification base does not lend the same support to his belief as it does in the actual world. Hence, the two conditions are not equivalent.

I submit that [Condition$_{\text{Justification-Base}}$] offers a fruitful reconstruction of the form of luck that is widely assumed to clash with knowledge. If, with respect to a person's justification base for the belief that *p*, it was a matter of luck that the belief turned out true, the person was lucky in a way that interferes with knowledge. There is hope that this notion can be cashed out in modal terms so as to match Pritchard's account of luck. Whether or not this makes for a definition of knowledge in terms of safe true belief is, of course, an entirely different matter.[23]

9. Conclusion

Let me summarize the main points. There is a form of benign luck that has gone unnoticed in the debate. Whether a set of beliefs plays a justificatory role in a context may be a matter of luck. This form of luck is compatible with knowledge. Moreover, luck with respect to a method of belief formation is also compatible with knowledge, as little reflection on [Barn County*] reveals. What is common to ways of belief formation and what I have labeled 'background information' is that both can play a justificatory role. Luck with respect to whatever can play a justificatory role for a subject in a situation is incompatible with knowledge. This seems intuitive: That justification or evidence and luck interact is already explicit in Engel's characterization of veritic luck. It appears that we can, by relativizing to a justification base of a subject with respect to a proposition in a context, offer a general interpretation of the relevant condition on knowledge in modal terms, without being committed to any particular view on how we should cash out justification, or support for a belief. Obviously, Gettier-cases and Russell's clock are covered: With respect to the justification base (including the method of belief-formation and available evidence), it is a matter of luck that the subject arrives at a

[23] See, for a critical discussion, Avram Hiller and Ram Neta, "Safety and Epistemic Luck," *Synthese* 158 (2007): 303-313.

true belief in Gettier scenarios; and with respect to the justification-base (observing a clock that has stopped working), it is a matter of luck that the subject forms a true belief about time.

Of course, there may be further constraints on knowledge. I did not intend to argue that based on the explication of luck proposed here, we arrive at a sufficient condition for knowledge.[24] Moreover, it is worth noting that having some sort of background-information or additional evidence regarding a belief that *p* need not always work as a remedy in cases where a belief is based on a bad method. As indicated above, one may want to hold that when a subject forms a true belief based on wishful thinking, or on a lucky guess, the subject does not acquire knowledge – independent of the additional evidence the subject may possess. In such cases, the subject was of course lucky to arrive at a truth. There is bad luck due to method alone. All I have argued is that the mere fact that the subject was lucky with respect to method alone is not sufficient to explain why the subject did not acquire knowledge (there may be types of methods, such as guesswork or wishful thinking, that do the trick – when using them, you will never acquire knowledge.) Finally, I did not intend to argue that a subject needs to *possess* background justification in order to acquire knowledge – hence, I did not touch upon questions pertaining to the internalism/externalism distinction.

Throughout this paper, I have tacitly assumed that in fact, there is a type of luck that is incompatible with knowledge acquisition, and that intuitions about scenarios are the guide to a successful characterization of luck and, possibly, a safety condition on knowledge. Recently, Baumann[25] has offered a number of scenarios in which, he suggests, we would ascribe knowledge to the subject, although the subject is, in a significant sense, lucky. Baumann's examples question, very roughly, the view that the feature of luck to undermine knowledge at *one* stage in a process relevant for belief formation or justification carries over to later stages. He considers, amongst other cases, causal chains where a watch with a reliable mechanism is set based on a Russellian clock (i.e. a clock that does not work but does, at a particular time of observation, 'indicate' the correct time) and is, much later, checked. Baumann suggests that subjects who, at later stages, form beliefs based on checking the watch can acquire knowledge. Early luck carries

[24] An interesting suggestion has recently been made by Schafer, who argues that based on considerations about knowledge ascriptions, we should offer a more general interpretation of the sort of luck that interferes with knowledge (Karl Schafer, "Knowledge and Two Forms of Non-Accidental Truth," *Philosophy and Phenomenological Research* LXXXIX (2014): 373-393). Given the fact that Schafer departs from a quite different perspective, I have ignored a discussion of his account in this paper.

[25] Baumann, "No Luck with Knowledge?"

over, but it gradually stops interfering with knowledge. The second set of examples includes an element of luck at some early stage in a chain of inferences, and he suggests that, again, at some point, the subject may have acquired knowledge based on these inferences.[26] Baumann then offers a more general diagnosis, suggesting that knowledge- and luck-ascriptions are contextual, in the sense that the relevant parameters will vary from context to context, so that 'absolutism'[27] about luck, or the luck-knowledge interaction, would turn out to be mistaken. Baumann also suggests that we should judge our account of knowledge (and luck) by its theoretical virtues. I have my sympathies for this take on the matter; we should aim at a fruitful explication of luck and knowledge. Up to some point, intuitions about scenarios may help, but they need not be regarded as being ultimately decisive.

Note that in this context, Baumann also suggests that Pritchard's and Engel's accounts fail because they are 'absolutist' in the sense that they are not flexible with respect to the kind of relativization (i.e. to evidence or method of belief formation). In this sense, the anti-luck condition discussed here would be absolutist as well. I think that this misrepresents the dialectical situation. One can consistently hold that the type of luck that interferes with knowledge is luck with respect to evidence, method, or justification base, and that this relativization is explanatory, and, at the same time, subscribe to some form of contextualism; whether or not luck of this sort does interfere with knowledge may still depend on additional contextual parameters. So, luck of this form may be present without interfering with knowledge. Whether it does may depend on contextual factors we cannot hold fixed once and for all. If, in a given context, luck of this sort interferes with knowledge, relativization to the justification base *explains why* it interferes with knowledge. Thus, absolutism, as opposed to contextualism, about luck is not a question of the relativization to method, evidence, or justification base. It is a matter of holding (or denying) the universal claim that luck of this type always, i.e. independent of further contextual conditions interferes with knowledge.

As a consequence, various other considerations may enter an assessment of the explication proposed here. Then, this paper should be regarded as an attempt to explicate *one* notion of epistemic luck that comes as close as we get to the notion epistemologists were typically aiming at. Based on intuitions alone, we should arrive at this explication. Further considerations, say, on theoretical

[26] Baumann also offers other types of scenarios, where luck stops interfering with knowledge due to other contextual conditions.

[27] Baumann, "No Luck with Knowledge?", 545.

elegance, suitability for a formal treatment etc., may suggest an alternative characterization.[28]

[28]**Acknowledgments:** I would like to thank Anna-Maria Eder and Insa Lawler for helpful discussions of an realier draft of this paper. Generous funding for this work was provided by the Volkswagen Foundation as part of the Dilthey-Fellowship "A Study in Explanatory Power," based at the University Duisburg-Essen.

SOLVING THE CURRENT GENERALITY PROBLEM

Kevin WALLBRIDGE

ABSTRACT: Many current popular views in epistemology require a belief to be the result of a reliable process (aka 'method of belief formation' or 'cognitive capacity') in order to count as knowledge. This means that the generality problem rears its head, i.e. the kind of process in question has to be spelt out, and this looks difficult to do without being either over or under-general. In response to this problem, I propose that we should adopt a more fine-grained account of the epistemic basing relation, at which point the generality problem becomes easy to solve.

KEYWORDS: basing relation, generality problem, process reliabilism

1. The Generality Problem

Despite the widely agreed failure of process reliabilism as a theory of justification, process reliabilism is alive and well in contemporary discussions of the theory of knowledge. Although epistemologists may not often refer to themselves as process reliabilists, all of the currently popular positions on the theory of knowledge, from modal epistemologies like safety and sensitivity theories, to virtue reliabilism, include versions of a process reliabilist condition (which is to say that they all consider it a *necessary* condition for a belief to be knowledge that it was formed as the result of a reliable process). They look to how a belief was formed (the relevant 'method of belief formation' as modal reliabilists say, or the belief forming 'capacities' or 'abilities' involved as virtue reliabilists say) and require that beliefs so-formed are reliable (generally modally reliable).

Given this, one well-known problem for traditional justification-centric process reliabilism, 'the generality problem,'[1] takes on a renewed urgency, since it also applies to many currently popular positions.[2] In a nutshell, the generality problem is the fact that an instance of belief formation can be described as the

[1] See Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," *Philosophical Studies* 89:1 (1998): 1-29.

[2] This is in addition to the argument made by Bishop, claiming that any plausible account of justification is going to have to face something like the generality problem. See Michael Bishop, "Why the Generality Problem is Everybody's Problem," *Philosophical Studies* 151:2 (2010): 285-298.

result of a number of different processes and the process reliabilist needs a principled way of determining which one exactly has to be reliable (in whatever sense of 'reliability' is at play) in order for their process reliability condition to be met. Furthermore, there is a potential trap in store for the unwitting process reliabilist: make the relevant processes too narrow and specific to particular occasions and the condition becomes too easily met (at the limit, the process will be so specific as to be a one-case, unrepeatable, event and so whether a belief is reliably formed will collapse into the question of whether it is actually true or not). On the other hand, make the process too broad and epistemically important features of the way that the belief was actually formed will end up being overlooked. For instance, if we count all the visually based beliefs that I form in fake-barn county in the same way then we have missed out something important, since we have rolled together unreliable beliefs about the presence of barns and perfectly reliable beliefs about the flashing 'check engine' light on my dashboard. The generality problem poses a challenge for process reliabilists then: to give a principled account of the relevant kind of process and to do so in such a way that it avoids the bind of being either too general *or* too specific.

When looking to assess the reliability of some belief, the generality problem demands that we specify which other (actual or counterfactual) cases of believing that p have to be accurate in order for the belief under consideration to count as reliable. It may seem that a modal reliability condition, like safety, provides an answer to this question: we need to look at just those beliefs which are formed in nearby possible worlds.[3] However, not *all* of the nearby worlds in which one believes that p are in fact relevant to the safety of one's belief. This is because the safety of a belief is relativised to the *method* by which it was formed. Roughly put, it does not matter if you might easily have falsely believed that p, so long as you wouldn't have done so by forming your belief in the same way as you actually do.

So although *only* nearby worlds are relevant to assessing the reliability of a belief according to safety, not *all* nearby worlds are (only those in which one forms the belief that p in the same way as one actually does). So the generality problem is not yet fully solved until we can give an account of the method of belief formation at play in a given case, and of course do so in a way which avoids the bind of being either too general or too specific.

[3] Of course, one may worry that this is too vague a notion to do the job required, but let us not dwell on this point.

2. Comesaña's Solution

Juan Comesaña takes up this challenge.[4] He suggests that the relevant kind of process (the one which needs to be reliable if your belief is to be knowledge)[5] is forming the belief that you did on *the basis of the evidence* that you did. I think that this focus on epistemic basing and reasons for belief is right: those are exactly the epistemically relevant things (they are exactly the kinds of things that matter for justification too) and so are exactly what we should be looking at. However Comesaña's suggestion involving very narrow processes is misguided; it falls into the trap of being too specific. Comesaña's suggestion is that:

> A belief is well-founded iff it is based on evidence E and "the type producing a belief that p based on evidence E is a reliable type."[6]

The problem with this is that if E refers to a very specific body of evidence, then the process is too narrow. For example, consider this twist on a fake barn case:

> Red barn[7]
>
> Henry is driving through fake barn county (an area populated with fake barn facades) and (twist 1) Henry *knows* that he is driving through fake barn county. *But* what he doesn't know is that (twist 2) the fake barns in fake barn county are always green, and the real ones always red. He sees a red barn in a field by the road and he believes on the basis of this visual evidence that there is a red barn in the field.

Clearly Henry's belief that there is a red barn isn't justified and it isn't knowledge. The problem is that his belief is formed in an unreliable way (e.g. it isn't safe or sensitive, and as he continues to form beliefs like this as he drives through fake barn county, many of them are false). But given that Henry's evidence E is his visual perception as of a red barn, and the belief that he forms on this basis is that there is a red barn, then Henry's belief forming process is reliable on Comesaña's account. After all, Henry's perceptions as of red barns *do* track the presence of red barns in the actual world as well as relevant counterfactual ones.

To avoid this problem, Comesaña's account needs to broaden the conception of the relevant evidence E, so that it refers not to a specific body of

---

[4] Juan Comesaña, "A Well-Founded Solution to the Generality Problem," *Philosophical Studies* 129:1 (2006): 27-47.

[5] Or, in the context that he approaches the challenge, justification.

[6] Juan Comesaña, "A Well-Founded Solution," 38.

[7] Kripke used a similar red barn case, but where Henry was not aware that he was in red barn country, and in relation to a point about knowledge. See Saul Kripke, *Philosophical Troubles: Collected Papers Vol I* (Oxford: Oxford University Press, 2011), 186.

evidence, but to a more general kind of evidence. Only in this way will the relevant method of belief formation be applicable to a suitably wide range of cases. He needs to give an account of what *kinds* of beliefs, formed on the basis of what *kind* of evidence, need to be reliable in order for a given belief count as knowledge. In other words, we have to say in which possible worlds does a belief that p count as being formed via the same method. But this is just an epicycle of the generality problem: the challenge remains to specify what kind of process is the relevant one, the one that has to be reliable.

3. A Novel Solution

My suggestion is that we should work with a more fine-grained conception of the roles that reasons play in belief formation, conceiving of them in a more psychologically plausible way. The usual method employed in epistemology involves attempting to capture the process of belief formation and the nature of reasons in terms of a picture on which a body of evidence stands in a causal and rationalising relation to a particular propositional belief. It seems to me that this is too coarse-grained a picture of what goes on to capture everything of epistemic interest. Once we have a more fine-grained alternative in view, the generality problem will not look so problematic.

Consider an example:

> In the morning you see £200 on the kitchen table and over breakfast Tina tells you that she is buying something off of a friend today. Later on you see Tina's friend Sam coming up to the house with an interestingly shaped box and a little while later you hear the sound of an electric guitar coming from Tina's room. On the basis of everything that you have seen and heard that day, you believe that Tina has bought an electric guitar from Sam for £200.

In this case, a simple picture accounting for what happens is that you have a certain body of perceptions as of certain events, this body of evidence constitutes a reason for you to form some belief, and so you form that belief. In other words, the sum of everything that you have seen and heard gives you sufficient reason to believe that Tina has bought an electric guitar from Sam for £200, and you form that belief on this basis. While this is doubtless true, this coarse-grained description fails to capture some of the important facts about the reasons you have for your believing as you do and the way that you form your belief as a result.

If you had not seen the £200 on the kitchen table then you would not have had reason to believe that Tina has bought an electric guitar from Sam for £200, you just would have had reason to believe that Tina has bought an electric guitar from Sam. Relatedly, if you had seen £400, not £200, on the kitchen table then

you would have had reason to believe that Tina has bought an electric guitar from Sam for £400, not £200. (One natural way of expressing what these considerations show is to say that your reason for believing that Tina bought a guitar from Sam *for £200* is your seeing that amount of money on the kitchen table.)

Reflection on this case seems to show that you have reason to believe as you do because various of the particular things that you have seen and heard give you reason to endorse different fine-grained aspects of the content of the proposition that Tina bought a guitar from Sam for £200. The reasons for which you believe are not all reasons for you to believe the whole of that particular proposition. They do not all give you some small degree of indiscriminate evidence for the whole proposition. Instead, each of the things that you have seen and heard gives you compelling evidence for some *particular aspect* of the content of the proposition: the who, the what, the how much, etc.

On this view, the epistemic reason-relation can relate quite specific features of one's body of evidence to fine-grained elements of the content of a given belief. More generally, this is an example of the fact that reasons are combinatorial and so the fact that some evidence base provides reason for a particular belief owes to the combinatorial effect of a complex of contributory reasons. (Which can be much more complicated than the toy example I have given.)

(The claim that some kind of combinatorial structure applies to the kinds of reasons that ordinary human thought engages with – even if the details do not exactly match up with the rough sketch that I have employed here, which has focused on a simple compositional case – can be argued for on the basis of the systematicity and productivity of our responsiveness to reasons. This indicates that combinatorial structure is at work in even seemingly simple cases like direct perceptual beliefs.)

With this alternative understanding of the structure of reasons as productively and systematically combinatorial, solving the generality problem becomes simple. Comesaña was along the right lines in thinking that process reliabilists should be focusing on the reasons which a subject is responding to in forming a particular belief. But we must then note that the reasons which a subject is responding to have a complex, combinatorial, structure. There is not just one kind of reason contributing to a given belief; there are many, all making individual contributions to *what* exactly there is overall reason to believe. Given this picture, a process reliabilist should maintain that *all* of these need to be reliable in order for a belief to be knowledge. This avoids both of the traps mentioned earlier: by requiring the reliability of *all* of these kinds of reasons responsiveness, nothing of epistemic importance is left out; but we do not consider

anything *too highly specific* since each of the individual kinds of reasons responsiveness involved can and would apply to many other beliefs as well and their overall reliability will therefore depend on their truth-conduciveness in these other cases too.

For instance, in the red barn case Henry believes that there is a red *barn* because of the barn-like quality of his perception, but this is not a reliable way of forming beliefs since it would easily lead him to have false beliefs about green barns.

4. Conclusion

The generality problem poses a question to process reliabilists: how should we understand the 'process' in process reliabilism; *what exactly* has to be reliable in order for a belief to be knowledge? I have argued that the answer we should give is that the *processes,* plural, relevant to this epistemic evaluation are *all* of the *many* kinds of reasons responsiveness which are involved in the formation of that belief, in accordance with an understanding of the structure of reasons as complex and combinatorial. This gives us a principled solution to the generality problem.

DISCUSSION NOTES/DEBATE

BEAT THE (BACKWARD) CLOCK[1]

Fred ADAMS, John A. BARKER, Murray CLARKE

ABSTRACT: In a recent very interesting and important challenge to tracking theories of knowledge, Williams & Sinhababu claim to have devised a counter-example to tracking theories of knowledge of a sort that escapes the defense of those theories by Adams & Clarke. In this paper we will explain why this is not true. Tracking theories are not undermined by the example of the backward clock, as interesting as the case is.

KEYWORDS: Dretske, backward clock, knowledge, Nozick, Sinhababu, tracking theories, Williams

I. Introduction

In a recent very interesting and important challenge to tracking theories of knowledge, Williams & Sinhababu[2] claim to have devised a counter-example to tracking theories of knowledge of a sort that escapes the defense of those theories by Adams & Clarke.[3] In this paper we will explain why this is not true. Tracking theories are not undermined by the example of the backward clock, as interesting as the case is.[4]

[1] We are very grateful to John Williams and Niel Sinhababu for helpful comments and discussion when Adams presented a version of our reply at Singapore Management University in March 2016. We are also grateful to Peter Baumann for useful suggestions and for pointing out similarities between the issues in the attack by Williams & Sinhababu and those discussed in Wolfgang Frietag, "Safety, Sensitivity and 'Distant' Epistemic Luck," Theoria 80 (2014): 44-61, and Fernando Broncano-Berrocal, "No Luck in the Distance: A Reply to Freitag," Theoria 82: 89-100.

[2] John Williams and Niel Sinhababu, "The Backward Clock, Truth-Tracking, and Safety," *Journal of Philosophy* 112, 1 (2015): 46-55.

[3] Fred Adams and Murray Clarke, "Resurrecting the Tracking Theories," *Australasian Journal of Philosophy* lxxxiii, 2 (2005): 207-21.

[4] Tracking theories are also not undermined by the examples of Tristan Haze as we've argued elsewhere (Fred Adams and Murray Clarke, "Two Non Counterexamples to Truth-Tracking Theories of Knowledge," *Logos & Episteme. An International Journal of Epistemology* VII, 1 (2016): 67-73). We think similar problems face both the attempted counter-examples by Williams & Sinhababu and Tristan Haze, "Two New Counterexamples to Truth-Tracking Theories of Knowledge," *Logos & Episteme. An International Journal of Epistemology* VI, 3 (2015): 309-11.

The counter-example of Williams & Sinhababu (hereafter W&S) is aimed at the sensitivity condition of tracking theories. Knowledge may be obtained, on tracking theories of knowledge, when p is true, one believes that p on the basis of a reason R or method M, and one's reason R (Dretske)[5] or method M (Nozick)[6] is *sensitive* to the truth – i.e. if p were not true one would not believe that p via R or M.[7] One would not believe the false p because one's reasons R or method M would be truth-tracking when knowledge-conducive and would not lead one astray. So for example, when you start your car and look to see if the oil pressure is fine, if the pressure gauge is working properly, it would not show the pressure normal unless it was normal. When the gauge satisfies this condition, it is sensitive to the truth about the pressure in the running engine. Your reason for believing the oil pressure is normal is that the gauge says it is normal. Your method of forming beliefs about the oil pressure in your running engine is reading the pressure gauge. So you know the pressure is normal because via this means you are tracking the truth. Thus, your method of forming your belief is the procedure you use to acquire your reason for believing, and your reason affords you knowledge only if it is sensitive to the truth – it wouldn't obtain unless your belief were true. This sensitivity condition is at the heart of tracking theories of knowledge. So it is here that W&S strike.

The plan of W&S is to devise an example that satisfies the conditions of tracking theories, the truth condition, the belief condition, and the sensitivity condition, but that still is not a case of knowledge. If their example were to succeed, it would show that tracking theories of knowledge are too weak because they permit cases of accidentally true (and sensitive) belief that meet the sufficiency conditions of the theory. As we say, we shall argue that they are not correct and that tracking theories of knowledge escape their attempt at finding a counter-example.

II. Normal Clock

W&S use a series of clock-examples to work up to their backward clock example. In their first example, a normal clock, all tracking conditions are met.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked

[5] Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1-22.
[6] Robert Nozick, *Philosophical Explanations* (Cambridge, M.A. Harvard University Press, 1981).
[7] We will not discuss Nozick' s *adherence* condition, as it is not relevant to the example of W&S. Our formulation of the sensitivity condition is designed to emphasize the similarities between the formulations of Dretske and Nozick.

> perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because the clock has continued to work perfectly reliably.

Here you believe the truth and your belief-forming method is 'looking at the clock,' or, more precisely, 'looking at the clock and determining what it says.' This method is sensitive to the truth, for the clock wouldn't say what is says if the correct time weren't 4:30 p.m. Hence, all conditions of the tracking theories are met and there is no counter-example.[8]

III. Stopped Clock

In the second example, going back to Russell,[9] they use the example of the stopped clock.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Like Normal Clock, it has an analogue design so its hands are supposed to sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because exactly twenty-four hours ago a stray fleck of dust chanced to enter the clock's mechanism, stopping it.

This is not offered as a counter-example to tracking theories. W&S hold that you do not know that it is 4:30 p.m., and we agree with this assessment. Since the clock has stopped, it no longer *says* anything, even though it continues to *display* '4:30.' In this case your method cannot consist in 'looking at the clock and determining what it *says,*' for it says nothing. Rather, your method consists in 'looking at the clock and determining what it *displays*,' and this method isn't sensitive to the truth – the clock would display '4:30' even if the correct time weren't 4:30. Of course, you do not know that the clock has stopped or you would not form a belief about the time on the basis of what it displays.

---

[8] Actually there is a bit of a problem here that will be important when we get to the backward-clock example. There is no second-hand. So one cannot use this clock to know when it is exactly 4:30. The best one could know is that it is approximately or nearly 4:30.

[9] Bertrand Russell, *Human Knowledge: Its Scope and Limits* (New York: Simon & Schuster, 1948).

IV. Backward Clock

This is the example that W&S claim is a counter-example to tracking theories of knowledge. They argue that it satisfies the conditions of the theory but that the subject does not have knowledge. So the theory must be too weak.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour. So your clock is designed to run perfectly reliably backwards during that hour. At 4 pm the hands of the clock jumped to 5 pm, and it has been running reliably backwards since then.

W&S believe that this example is a counter-example because they believe it fits all of the conditions of tracking theories. The belief that it is 4:30 p.m. is true. It is believed true. And the belief is sensitive to the truth because, given the way this clock works, even though it is running backwards from 4:00 p.m. to 5:00 p.m., since it starts running backward from '5:00' (at 4:00 p.m.), by 4:30 p.m. the clock will say it is '4:30' p.m. And given the way the clock works, it wouldn't say '4:30' unless it was 4:30.[10]

Now there are several reasons why we maintain that this is not a counter-example to tracking theories of knowledge. We shall now go through those reasons.

V. Not a Counter-Example

First, there is no second-hand on the clock. We suspect that they designed all the examples this way because they thought that if a subject observed the backward clock second-hand going backwards, the subject might not trust the clock. The subject might not believe what the clock says. But since the subject cannot see when the second-hand hits '12,' the subject could never use this clock to know that it is exactly 4:30 p.m. Hence, this is a serious defect in their example. At best one could know that it is approximately or nearly 4:30 p.m. (perhaps within one minute). But we agree with W&S that one will not even know this by using this clock.

Second, W&S say that the reason one lacks knowledge in the backward-clock example is: "We claim that Backward Clock is not a case of knowledge because (as in Stopped Clock) your belief that it is 4:30 pm is luckily true." We

[10] Actually, it would. It would display '4:30' twice a day. So it would display it was 4:30 (p.m.) even if it were 4:30 a.m. It is totally ambiguous between a.m. and p.m. when it displays '4:30.'

will take this to mean that the belief is accidentally true. Now if true, where does the accident occur? There are two possibilities. It could happen between the time and the reading on the clock. Or it could happen between the reading on the clock and the truth of the belief of the believer. Actually, we believe there is accident on both sides of this causal chain, as we will now explain.

Our third point concerns W&S's description of the clock: "Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour. So your clock is designed to run perfectly reliably backwards during that hour." This rather vague description of the clock lends itself to at least two interpretations. The most plausible interpretation is that the clock was designed to deceive viewers during the cursed hour – the clock was, so to speak, designed to say something false, i.e., to *lie*, about the time during this hour. A second, less plausible, interpretation is that the clock was designed to be used by cult members as a 'countdown' clock for the period between 4:00 and 5:00. The clock enabled members to ascertain the correct time during the cursed hour, but wasn't intended to be viewed by the general public, and wasn't designed to deceive anyone. Perhaps cult members would have been chagrined to learn that one of their special clocks was in the possession of someone who didn't belong to the cult.

Before focusing on the first interpretation, we pause to note that on the second interpretation, the Backward Clock case is basically a 'non-starter.' The ordinary person, call her Betty, would be unable to determine what the clock is saying during the hour between 4:00 and 5:00. When the clock displays '4:35,' it is saying that the correct time is 4:25, and when it displays '4:25,' it is saying that the correct time is 4:35. It so happens that when the clock displays '4:30,' it is saying that the correct time is 4:30, but Betty would have no way of telling that this is what it is saying. She would form the belief that it is 4:30, but in virtue of the fact that she wouldn't understand what the clock is saying, it would be merely a chronometric accident or a coincidence that this belief would be true.

In the remainder of this paper we'll focus on the most plausible interpretation, that is, that the clock was designed by the cult clockmakers to deceive viewers during the cursed hour. Unlike an ordinary clock, which is designed to say what the correct time is, the cult clock was designed to say something false about the time, i.e., to 'lie' about the time. The scheme that the lying clockmaker – call him Ted – adopted to achieve this aim was making the clock run backwards from 4:00 to 5:00. But his scheme was flawed, for it failed to say something false about the time at 4:30. He could easily have eliminated this flaw, say, by making the clock run more slowly to ensure that it wouldn't display

the correct time at any point during the hour. Granted, Ted wasn't a perfect liar. But his mistake wouldn't enable you to learn the correct time at 4:30. As Dretske pointed out, "If your reasons for believing P are such that you might have them when P is false, then they aren't good enough to know that P is true. You need something more... That is why you can't learn – can't come to know – that P is true if all you have to go on is the word of a person who might lie about whether or not P is so. This is just another way of saying that knowledge requires reasons or evidence (in this case, testimony) you wouldn't have if what you end up believing were false. You can learn things from people, yes, but only from people who wouldn't say it unless it were true."[11] The clock in the Normal Clock case wouldn't have said that the time was 4:30 by displaying '4:30' if it hadn't been 4:30. Ted's clock, however, might have done this even if it hadn't been 4:30.

Fourth, there is an accidental connection between the clock display and the belief of the subject awakening from the nap. For any time other than exactly 4:30, the subject's belief during that hour-long period will be false. Why? Because the clock lies for all but one moment during that hour-long period. And worst of all, there is nothing in the signal sent by the clock to differentiate when it is telling the false time from when it is telling a true time.

This should remind one of the "little boy who cried 'wolf.'" The boy cries 'wolf' over and over when there is no wolf. Then on the one occasion when there is a wolf and he cries 'wolf,' his cry has become so equivocal, no one can tell from his cry that a wolf is actually there on that one occasion. His cry of 'wolf' still means wolf, but it does not carry the information that there is a wolf.[12] Similarly, the clock's face emits false testimony for 59 minutes during that hour from 4:00 to 5:00. There is no way that one could tell from this equivocal messenger which clock display (if any) expresses a truth rather than a falsity.

Now W&S will no doubt insist "but at 4:30, it is still true that it would not display '4:30' unless it were actually 4:30." Consider the following. Suppose the boy who cried 'wolf' figured out that if he keeps saying 'wolf' when there is no wolf, then if ever there actually is a wolf, no one will come when he cries 'wolf.' So after his last prank cry he adopts the new policy that he won't cry wolf again unless there actually is a wolf. Then the wolf shows and he cries 'wolf.'

The problem is that although the conditions have now changed and now, as opposed to before, he would not cry 'wolf' unless there was a wolf, no one hearing

[11] Fred Dretske, "Reply to Hawthorne," in *Contemporary Debates in Epistemology*, ed. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 43-46.

[12] Fred Dretske, *Knowledge and the Flow of Information* (Cambridge, M.A.: MIT/Bradford, 1981).

his cry could discriminate the change of circumstances. His new cry will be just as equivocal now as it was before. Perhaps if his cries changed in tone or pitch or frequency, this would carry information that it was *for real*, but barring that there would be no way to tell.

On the clock face there is nothing to distinguish its '4:30' display from the other displays between 4:00 and 5:00. So even if it is true that it would not display '4:30' unless it were 4:30, there is no way on earth for a naïve subject viewing its face to discriminate the difference of contexts. The clock-viewer is getting just as equivocal a message from the clock as those hearing the little boy cry 'wolf' *after his decision not to cry unless there really is a wolf.* So even if a shepherd for some reason only heard the cry after the boy changed his *modus operandi*, the shepherd would not know there was a wolf precisely because there was nothing in the signal to differentiate the change in contexts (that the boy changed his *modus operandi*).

So, returning to the clock, while the person waking from the nap happens to acquire a true belief that it is 4:30, the clock's display is equivocal even if the clock wouldn't display '4:30' unless the time were 4:30. No naïve reader could differentiate the true from the false clock displays during that hour. So if the subject acquires a true belief it is doubly lucky/accidental. For there is luck/accident in the link from the actual time to the clock's displaying 4:30, and there is luck/accident in the link from the clock's displaying 4:30 to the subject's belief. Hence, the reason the subject lacks knowledge is that neither the clock nor his belief is tracking the truth. His belief is that it is 4:30, and it happens to be 4:30. But it is not the case that he believes it is 4:30 *because* it is 4:30 – his believing it to be 4:30 is not explained by the fact that it is 4:30. For all 'X' between '4:00' and '5:00,' it is false the clock would not display 'X' unless the time were X. So the method (or reason) that gives rise to the subject's belief is not a truth-tracking method (or reason), for it is too equivocal to yield knowledge.

VI. Possible Reply by Williams & Sinhababu

When Adams & Clarke[13] defended tracking theories of knowledge, we pointed out that the sensitivity condition is relativized to a reason R or a belief-forming method M and these are also relativized to one's environmental circumstances C. So a functioning compass can tell you where geomagnetic north is, but not down in a mineshaft. A mercury thermometer can tell you your child's temperature, but

[13] Fred Adams and Murray Clarke, "Resurrecting the Tracking Theories," *Australasian Journal of Philosophy* 83, 2 (2005): 207-221.

only if you shake it down before re-use. Reasons and methods work in some circumstances and under some conditions, but not under others.

In our defense of tracking theories against the example of the backward clock, we have presupposed that one's method for forming the belief that it is 4:30 is looking at the clock and determining what it says, and one's reason for believing that it is 4:30 is the clock's saying that it is 4:30.

Now we suspect that W&S will respond that the method was not that. They may insist that the method was 'looking at the clock at 4:30.'[14] The reason the nap-taker believed it was 4:30 p.m. is that he looked at the clock at 4:30 p.m. However, we maintain that this may be the cause of the nap-taker's belief, but it is hardly the evidential reason why he believes it is 4:30.[15]

First, the only way he comes to believe it is 4:30 is by looking at the clock and determining what it says. He evidentially does not believe it is 4:30 by looking at the clock at 4:30 and determining what it says. That is evidentially circular. If he knew to look at the clock at 4:30, he would not need the clock.

Second, if one asked, "why did you believe it was 4:30?" he would answer "because the clock said so." He would not answer, "because at 4:30 the clock said so." Clearly, his reason or method of belief-formation was "looking at the clock, and determining what it said, period," and as we have pointed out, that method is not a truth-tracking method in the backward-clock case. Using that method, it is sheer accident/luck that one's belief comes out true, even at 4:30. This is why tracking theories would say that one does not know in the case of the backward clock, and it is why the example is not a counter-example to tracking theories.

VII. Conclusion

We have examined the purported counter-example to tracking theories offered by W&S. The backward-clock example is clever, interesting, and important. We have argued that the example is not a counter-example because on tracking theories themselves the belief is too accidentally true to count as an example of knowledge. We have explained why the sensitivity condition is not met in the example, despite the arguments of W&S to the contrary. The belief-forming method in the

[14] When Adams presented our reply to them in person in March 2016, they gave this response.

[15] Independently, we realized that if one restricts the method to reading the clock at precisely 4:30 in order to make the counterfactual true, then W&S would appear to have trouble rejecting that one knows in the case of stopped clock. Peter Baumann in correspondence confirmed our suspicion saying: "W&S would have to argue, by parity of reasons, that the subject in the stopped clock case also knows the time according to Sensitivity (don't they use the method of looking at the clock at THAT time?). But this would be based on a caricature of Sensitivity."

example is not a truth-tracking method. The evidence upon which the belief is formed is equivocal. There is not enough information in the evidence to differentiate the true from the false clock readings. One's belief is not formed on the basis of a sensitive method or reason – hence one's belief (though true) does not track the truth.[16] For this reason, one does not know that it is 4:30 p.m. and tracking theories of knowledge agree that one does not know. Thus, there is no counter-example.

[16] Again we thank Peter Baumann for pointing out: "It seems to me that W&S are working with a conception of belief-sensitivity (this belief would not be tokened if it were false) while you're working with a conception of method-sensitivity (no (or almost no, etc.) beliefs from set X would be tokened if false). I think Sosa and Pritchard have, under pressure from counter-examples, made a similar move for their safety theories."

THERE'S NOTHING TO BEAT A BACKWARD CLOCK: A REJOINDER TO ADAMS, BARKER AND CLARKE

John N. WILLIAMS

ABSTRACT: Neil Sinhababu and I presented *Backward Clock*, an original counterexample to Robert Nozick's truth-tracking analysis of propositional knowledge. Fred Adams, John Barker and Murray Clarke argue that *Backward Clock* is no such counterexample. Their argument fails to nullify *Backward Clock* which also shows that other tracking analyses, such as Dretske's and one that Adams *et al.* may well have in mind, are inadequate.

KEYWORDS: Dretske, methods, Nozick, sensitive belief, truth-tracking

In "The Backward Clock, Truth-Tracking, and Safety," Neil Sinhababu and I presented *Backward Clock*, an original counterexample to Robert Nozick's truth-tracking analysis of propositional knowledge.[1] In "Beat the (Backward) Clock," Fred Adams, John Barker and Murray Clarke argue that *Backward Clock* is no such counterexample.[2] Their argument fails to nullify *Backward Clock* which also shows that other tracking analyses, such as Dretske's and one that Adams, Barker and Clarke may well have in mind, are inadequate. When what counts is derailing tracking analyses, there's nothing to beat a backward clock like ours.

1. Nozick's Analysis of Knowledge and the Backward Clock

Among truth-tracking analyses of knowledge we sought to fault only *Nozick's analysis of knowledge*, which we formulated as follows.

S knows that *p*, using method *M* of arriving at a belief whether *p*, just in case

(1) *p*

(2) *S* believes, using *M*, that *p.*

[1] John N. Williams and Neil Sinhababu, "The Backward Clock, Truth-Tracking, and Safety," *Journal of Philosophy* 112, 1 (2015): 46–55.

[2] Fred Adams, John A. Barker and Murray Clarke, "Beat the (Backward) Clock," *Logos & Episteme. An International Journal of Epistemology* VII, 3 (2016): 353-361

(3) In the closest (that is, most similar) worlds to the actual world in which not-*p* (and in which *S* uses *M*), *S* does not believe that *p*.

(4) In the closest (that is, most similar) worlds to the actual world in which *p* (and in which *S* uses *M*), *S* believes that *p*.[3]

There are a couple of things we should note about this analysis. (3) is commonly known as the 'sensitivity condition,' meaning that *S*'s belief that *p* is sensitive to falsehood; roughly, she would not have that belief if it were false. (4) is commonly known as the 'adherence condition,' meaning that *S*'s belief that *p* adheres to the truth; roughly, were she to have that belief in slightly changed circumstances, then it would still be true. A belief that is both sensitive to falsehood and adherent to truth is said to be 'truth-tracking.' The analysis proceeds in terms of a method and requires that (3) invoke the method that *S* actually uses (as mentioned in the *definiendum*) to arrive at her belief that *p*. Nozick introduces methods into his analysis, not as a way of elucidating sensitivity, but in order to avoid a counterexample.[4] We followed Nozick in taking sensitivity (and indeed adherence and truth-tracking) to be a condition on *S*'s *belief* that *p*, not on her *method M* of arriving at that belief.[5] We argued that *S*

[3] Williams and Sinhababu,"The Backward Clock," 46. Our formulation is faithful to Nozick, although it is not *verbatim*. In *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 179, he says

> Let us define a technical locution, S knows, via method (or way of believing) M , that *p*:
>
> (1) *p* is true.
>
> (2) S believes, via method or way of coming to believe M, that *p*.
>
> (3) If *p* weren't true and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*.
>
> (4) If *p* were true and S were to use M to arrive at a belief whether (or not) *p*, then S would believe, via M, that *p*.

Although this formulation does not explicitly mention possible worlds, Nozick is clear that his subjunctives (3) and (4) can be expressed as ours and announces that he will sometimes use them that way (*Philosophical Explanations*, 173–174).

[4] Nozick notes that (3) must be formulated in terms of the method that *S* actually uses, in order to avoid *Grandmother*. A grandmother sees her grandson is well when he comes to visit but if he were too unwell to visit, then relatives would tell her that he is well to spare her upset. She arrives at the true belief that he is well via the method of looking at him, yet if he were unwell then she would still believe that he is well via the *different* method of testimony. So (3) is false, but nonetheless she knows that he is well (*Philosophical Explanations*, 179).

[5] Nozick says of (3) that "it tells us how his belief state is sensitive to the truth-value of *p*. It tells us how his *belief state* is sensitive to *p*'s falsity, but not how it is sensitive to *p*'s truth" (*Philosophical Explanations*, 176, my italics). We called a belief that satisfies (3), 'sensitive to falsehood' or just 'sensitive' (Williams and Sinhababu, "The Backward Clock," 47–48, 51–52 and 55) following common parlance among commentators (for example, Tim Black and Peter Murphy, "In Defense of Sensitivity," *Synthese* 154, 1 (2007): 53–71, 54 and 58, Rachael Briggs

does not know that *p* in *Backward Clock*, but that this example satisfies (1)-(4), thus showing that Nozick's analysis, as given above, is too weak, predicting knowledge where there is ignorance. In order to support this claim, we first gave two other examples, *Normal Clock* and *Stopped Clock*.

We described *Normal Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because the clock has continued to work perfectly reliably.[6]

There is the oddity of the missing second hand. We left it off to ensure parity with *Stopped Clock* and *Backward Clock*, to be described shortly. Let us postpone examining this oddity until the next section. Your true belief that it is 4:30 pm is sensitive to falsehood. Had it been any time other than 4:30 pm when you looked at the clock, then you would not believe that it is 4:30 pm. Your true belief that it is 4:30 pm is also truth-adherent. Had you looked at the clock at 4:30

---

and Daniel Nolan, "Mad, Bad and Dangerous to Know," *Analysis* 72, 2 (2012): 314–16, Keith DeRose, "Solving the Skeptical Problem," in *Skepticism: A Contemporary Reader*, ed. Keith DeRose and Ted A. Warfield (New York: Oxford University Press, 1999), 197, Lars Gundersen, "Tracking, Epistemic Dispositions and the Conditional Analysis," *Erkenntnis* 72, 3 (2010): 358, David Manley, "Safety, Content, Apriority, Self-Knowledge," *Journal of Philosophy* 104, 8, (2007): 419 and Duncan Pritchard, "Anti-Luck Virtue Epistemology," *Journal of Philosophy* 109, 3 (2012): 247–79 and 250–254. For Nozick, a belief that is sensitive to truth is what he later calls 'adherent' (*Philosophical Explanations*, 212) or as we put it, 'truth-adherent' (Williams and Sinhababu, "The Backward Clock," 49, 51 and 55). Nozick tells us that "Perfect sensitivity would involve *beliefs* and facts varying together" (*Philosophical Explanations*, 176, my italics). He then says "Let us say of a person who believes that *p*, which is true, that when 3 and 4 hold, his *belief* tracks the truth that *p*" (*Philosophical Explanations*, 17, my italics). So against Tristan Haze ("Reply to Adams and Clarke," *Logos & Episteme. An International Journal of Epistemology* VII, 2 (2016): 224, note 11) talk of 'tracking the truth,' does appear in Nozick's official account, although not in his analysis. Haze independently hits the nail on the head by observing that "we cannot properly say: the account requires that the method M tracks the truth in general. That is a mischaracterization of Nozick's theory" (Haze, "Reply to Adams and Clarke," 224, note 11). The conjunction of 3 and 4 holding is also what commentators mean by a 'truth-tracking' belief (for example Gundersen, "Tracking," 358 and Raymond Martin, "Tracking Nozick's Sceptic: A Better Method," *Analysis* 43, 1 (1983): 28–33). Notice that for Nozick, as for us and his other commentators, it is *beliefs* that are sensitive to falsehood, not methods or ways. For Nozick, as for us and his other commentators, it is *beliefs* that are truth-adherent or truth-tracking, not methods or ways. This will become important in section 6.

[6] Williams and Sinhababu, "The Backward Clock," 46–47.

pm while being slightly closer to it, then you would still believe that it is 4:30 pm. So far so good for Nozick's analysis, because surely you do know that it is 4:30 pm. Adams, Barker and Clarke do not contest this.[7]

Then we described *Stopped Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Like *Normal Clock*, it has an analogue design so its hands are supposed to sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because exactly twenty-four hours ago a stray fleck of dust chanced to enter the clock's mechanism, stopping it.[8]

Your belief that it is 4:30 pm is insensitive to falsehood. If it were not 4:30 pm but some other time, then by looking at the clock you would still believe – but then falsely – that it is 4:30 pm. This is more good news for Nozick's analysis, since surely you do not know that it is 4:30 pm. One very plausible explanation of your ignorance is that your belief is luckily true. You were lucky to look at the clock exactly twenty-four hours after it stopped working, at the only instant during the hour when you nap at which its hands could have pointed to the correct time. Adams, Barker and Clarke do not contest this either.[9]

Finally, we described *Backward Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour. So your clock is designed to run perfectly reliably backwards during that hour. At 4 pm the hands of the clock jumped to 5 pm, and it has been running reliably backwards since then. This clock is analogue so its hands sweep its face continuously, but it has no second hand so you cannot tell that it is running backwards from a quick glance.

[7] They say that "all conditions of the tracking theories are met and there is no counter-example" (Adams, Barker and Clarke, "Beat the Clock," 355). By the plural 'theories' they mean Nozick's and Fred Dretske's early 1971 theory, to be discussed in section 5. We nowhere mentioned Dretske. More importantly, their understanding of Nozick's theory does not coincide with ours or any of the commentators I mention in note 5. In fact the tracking theory they defend is not Nozick's, as I will show in sections 1 and 6.

[8] Williams and Sinhababu, "The Backward Clock," 47.

[9] Of our observation that you do not know that it is 4:30 pm in *Stopped Clock*, they say that "we agree with this assessment" (Adams, Barker and Clarke, "Beat the Clock," 355).

> Awaking, you look at the clock at exactly 4:30 pm and observe that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm.[10]

As in *Stopped Clock*, your true belief that it is 4:30 pm is luckily true, for in both cases you were lucky to look at it at exactly 4:30 pm, at the only instant during the hour when you nap at which its hands could have pointed to the correct time. Thus in both cases you do not know that it is 4:30 pm. Your belief is also truth-adherent. Had you looked at the clock at 4:30 pm while being slightly closer to it, then you would still believe that it is 4:30 pm. In other words, in worlds close to the actual world in which it is 4:30 pm (and in which you look at your clock to tell the time), you believe that it is 4:30 pm.

But unlike in *Stopped Clock*, your belief that it is 4:30 pm is sensitive to falsehood, or in other words, satisfies (3). If it were not 4:30 pm but some other time, then by looking at the clock you would not believe that it is 4:30 pm. Instead you would form *some other false belief* about what time it is. For example, if you had looked at it at 4:31 pm, then you would not form the false belief that it is 4:30 pm. Instead you would form the false belief that it is 4:29 pm. To satisfy (3), you need not form a true belief about what time it is in the counterfactual situation (as you do in *Normal Clock*). You only need to fail to form a particular false belief – perhaps, by forming a different false belief about what time it is instead (as you do in *Backward Clock*). It is worth noting that *Backward Clock* can be seen to satisfy (3) without using the term 'sensitive:' in worlds close to the actual world in which it is not 4:30 pm but, say, 4:31 pm, and in which you look at your clock to tell the time, you do not believe that it is 4:30 pm. Instead you believe that it is 4:29 pm. Thus Nozick's analysis is too weak, predicting knowledge where there is none.

It is far from obvious that Adams, Barker and Clarke contest any of this either. While they tell us that our "claim to have devised a counter-example … is not true,"[11] they do not dispute the fact that your belief that it is 4:30 pm is truth-adherent, saying that that "Nozick's *adherence* condition … is not relevant to the example."[12] They do not dispute the fact that you have a true belief that it is 4:30 pm, saying that "the person awaking from the nap happens to acquire a true belief that it is 4:30."[13] They do not dispute the fact that you do not know that it is 4:30 pm, saying that "the *reason that the subject lacks knowledge* is that neither the clock nor his belief is tracking the truth."[14] So if they contest anything, they must

[10] Williams and Sinhababu, "The Backward Clock," 48.
[11] Adams, Barker and Clarke, "Beat the Clock," 353.
[12] Adams, Barker and Clarke, "Beat the Clock," 354, note 7.
[13] Adams, Barker and Clarke, "Beat the Clock," 359.
[14] Adams, Barker and Clarke, "Beat the Clock," 359, my italics.

dispute the fact that (3) is true of your belief that it is 4:30 pm. They do not dispute this in so many words. The nearest they come to disputing it is in the following passage.

> Now W&S will no doubt insist "but at 4:30 it is still true that it would not display '4:30' unless it were actually 4:30". …the clock's display is equivocal even if the clock wouldn't display '4:30' unless the time were 4:30. No naïve reader could differentiate the true from the false clock displays during that hour.[15]

This might make us suspect that Adams, Barker and Clarke are defending a different tracking theory from Nozick's. This is further confirmed by the fact that they argue that you do not have a truth-tracking *method* of forming the belief that it is 4:30 pm.[16] But Nozick's analysis is not elucidated in terms of a truth-tracking *method*, but in terms of a truth-tracking *belief*. I will return to this point in section 6. First however, let us examine their objections (but not quite in the order they give them).

2. No Second Hand on the Clocks? – No Problem

There is no second hand on any of the three clocks. Adams, Barker and Clarke comment that

> We suspect that they designed all the examples this way because they thought that if a subject observed the backward clock second-hand going backwards, the subject might not trust the clock.[17]

This is correct provided 'trust' is read as 'believe.' In fact, one reason why we left the second hand off *Backward Clock* is because in a realistic case in which you look at where its hands are pointing, you would recognize that it is running backwards. But then you would not believe that it is 4:30 pm, so we would have no counterexample.[18] We wanted to maximize parity among the three clocks. So we left the second hand off *Normal Clock* and *Stopped Clock* as well.

---

[15] Adams, Barker and Clarke, "Beat the Clock," 358-359.

[16] For example, "So the method (or reason) that gives rise to the subject's belief is not a truth-tracking *method* (or reason), for it is too equivocal to yield knowledge." (Adams, Barker and Clarke, "Beat the Clock," 359, my italics)

[17] Adams, Barker and Clarke, "Beat the Clock," 356.

[18] There is a second reason as well. Relatedly, we anticipated an objection as follows

> It might also be claimed that in *Backward Clock* you are not justified in forming any belief about what time it is by looking at the clock during its backward-running hour, because to be so justified you would have to check that its hands are still moving forwards (Williams and Sinhababu, "The Backward Clock," 50).

Adams, Barker and Clarke now object that

> since the subject cannot see when the second-hand hits '12,' the subject could never use this clock to know that it is exactly 4:30 p.m. Hence, this is a serious defect in their example.[19]

This objection is misstated. It cannot be that in *Backward Clock* you cannot *know* that it is 4:30 pm. This is true (because your belief that it is 4:30 pm is luckily true) yet (1)-(4) are all true. That putative objection is a vindication! More charitably, Adams, Barker and Clarke might mean that you wouldn't acquire the *belief* that it is 4:30 pm, because that would be the belief that it is *exactly* 4:30 pm, and you wouldn't come to *that* belief by looking at the position of the hands when you wake. This robs us of a counterexample.

I have two responses to this objection. Firstly, you might indeed come to believe that it is exactly 4:30 pm by observing the position of the hands when you wake. We may stipulate (as with *Normal Clock* and *Stopped Clock*) that you are close enough to the clock to see exactly where its hour and minute hands are pointing, and as you glance at it, you observe that its minute hand points exactly at the numeral 6 and its hour hand points exactly at the point equidistant in the arc from the numeral 4 to the numeral 5. Alternatively, we could stipulate that *Normal Clock* and *Stopped Clock* have a second hand that ticks forwards in discreet one-second jumps. In the hour that you nap, *Backward Clock*'s second hand ticks backwards in discreet one-second jumps. In each case you wake and observe the position of the hands of your clock. In *Normal Clock* and *Backward Clock* you observe the second hand of your clock at the instant it has ticked to the numeral 12 and then look away before it ticks past it.

Secondly, suppose (counterfactually) that the absence of the second hand *does* mean that you can only know that it is approximately 4:30 pm. Let us stipulate that you take 'approximately *t*' to mean 'within the period from (and including) *t* minus one minute to (and including) *t* plus one minute.' On this stipulation, approximate times include but do not exhaust exact times. Also suppose that you are cautious enough to form beliefs only about approximate times. In *Backward Clock* you wake and at exactly 4:29 pm observe the hands of your clock pointing to approximately 4:30 pm, so that its hour hand points to somewhere close to the point equidistant in the arc from the numerals 4 to 5 and its minute hand points to somewhere very close to the numeral 6. Accordingly,

---

We thought that we would be more saliently vulnerable to that objection had we kept the second hand on *Backward Clock.* But even if you are not justified in believing that it is not 4:30 pm, this will not save Nozick's analysis, because it does not mention justification.

[19]Adams, Barker and Clarke, "Beat the Clock," 356.

you form the true belief that it is approximately 4:30 pm. You do not know this because your belief is luckily true. You were lucky to look at it at between 4:29 pm and 4:31 pm, this being the only period of the hour when you nap during which its hands could have pointed to the approximately correct time. But now (3) is true. If it were not approximately 4:30 pm but say, exactly 4:32 pm, then you wouldn't believe that it was approximately 4:30 pm. Instead you would believe that it is approximately 4:28 pm. Once again Nozick's analysis predicts knowledge where there is ignorance.

3. The Irrelevant Intentions of the Cult

Adams, Barker and Clarke single out our explanation of why *Backward Clock* runs perfectly reliably backwards from 5 pm to 4 pm. This was that "your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour."[20] They say that this may be given one of two 'interpretations.'[21] The first of these is that the cult intends to deceive you into holding false beliefs about the time, or as they put it, "to say something false, i.e. to *lie*, about the time."[22] The second is that it intends it for its own use to tell the time during the cursed hour. They then argue that on either interpretation we are left with no counterexample.

Of the second interpretation (this would be more accurately called a speculation about the cult's intentions), they write that

> The ordinary person, call her Betty, would be unable to determine what the clock is saying during the hour between 4:00 and 5:00. When the clock displays '4:35,' it is saying that the correct time is 4:25, and when it displays '4:25,' it is saying that the correct time is 4:35. It so happens that when the clock displays '4:30,' it is saying that the correct time is 4:30, but Betty would have no way of telling that this is what it is saying.[23]

We nowhere use the metaphor 'say' in any of our discussion of any of the clocks. Adams, Barker and Clarke appear to think that for you to know what the clock 'says' is to know how to use your observations of the positions of its hands in order to know what time it is. This means that you must know not only what positions these are but also how such positions are intended to represent the time. Of course the cult, but not you, knows that. If you were to look at the clock at 4:25 pm, you would think that the position of its hands are intended to represent

[20] Williams and Sinhababu, "The Backward Clock," 48.
[21] Adams, Barker and Clarke, "Beat the Clock," 357.
[22] Adams, Barker and Clarke, "Beat the Clock," 357.
[23] Adams, Barker and Clarke, "Beat the Clock," 357.

the time as being 4:25 pm, whereas they are intended to represent the time as being 4:35 pm. But when you look at it at 4:30 pm, the cult does indeed intend the position of its hands to represent the time as being 4:30 pm. After all, they designed the clock to run perfectly reliably backwards from 5:00 pm to 4:00 pm. So now your understanding of what the position of the hands represents coincides exactly with what the cult intends, although you don't know that. But then Adams, Barker and Clarke immediately conclude that

> She would form the belief that it is 4:30, but in virtue of the fact that she wouldn't understand what the clock is saying, it would be merely a chronometric *accident or a coincidence* that this belief would be true.[24]

I agree, but I fail to see how this deprives us of a counterexample to Nozick's analysis. Your (or her) belief that it is 4:30 pm is coincidentally or luckily true (the time that you look at your clock coinciding with the only time during the hour that you nap that its hands could have pointed to the correct time), with the result that you do not know that it is 4:30 pm. But (3) is true. To repeat, in worlds close to the actual world in which it is not 4:30 pm (but, say, 4:31 pm) and in which you look at your clock to tell the time, you do not believe that it is 4:30 pm. Instead you believe that it is 4:29 pm.

Of the first interpretation of the cult's intentions, that it intends to 'lie' to you about the time, Adams, Barker and Clarke say that this "scheme was flawed, for it failed to say something false about the time at 4:30."[25] They continue that this

> … mistake wouldn't enable you to learn the correct time at 4:30. As Dretske pointed out, "If your reasons for believing P are such that you might have them when P is false, then they aren't good enough to know that P is true …"[26]

At this point Adams, Barker and Clarke appear to defend Dretske's early analysis of knowledge, not Nozick's, because (1)-(4) do not mention reasons, let alone incorporate Dretske's conditional. So the best that they may claim is that Dretske's analysis, rather than Nozick's, survives *Backward Clock*. I will show in section 5 that it doesn't survive it either.

However I agree that you couldn't learn that it is 4:30 pm, because whatever you learn you know and you can't know that it is 4:30 pm because your belief that it is 4:30 pm is luckily true. Once again our counterexample to Nozick's analysis is unscathed, since (1)-(4) remain true.

---

[24]Adams, Barker and Clarke, "Beat the Clock," 357, my italics.

[25]Adams, Barker and Clarke, "Beat the Clock," 357.

[26]Adams, Barker and Clarke, "Beat the Clock," 358.

Adams, Barker and Clarke conclude, naming the cult's clock designer 'Ted', that

> The clock in the Normal Clock case wouldn't have said that the time was 4:30 by displaying '4:30' if it hadn't been 4:30. Ted's clock, however, might have done this even if it hadn't been 4:30.[27]

This last claim is simply false. We stipulated that in the actual world, the clock runs perfectly reliably backwards from 5:00 pm to 4:00 pm. So the only time at which its hands can point to 4:30 pm is when it is 4:30 pm. Adams, Barker and Clarke point out that the cult could design the clock so its hands wouldn't point to the correct time at any time during the hour that you nap (say, by making it run backwards more slowly).[28] Perhaps they had that possibility in mind. But as we described *Backward Clock*, worlds close to the *actual* circumstances in which you look at it cannot include those in which its *mechanism* differs from that which makes it run perfectly reliably backwards from 5:00 pm to 4:00 pm. As we said, this is because the truth-adherence of your belief that it is 4:30 pm in *Normal Clock* resides in the fact that you would still have that belief in slightly changed circumstances in which the mechanism of the clock continues to work perfectly reliably. Likewise, the worlds close to the actual circumstances of *Stopped Clock* surely include those in which the mechanism of the clock is stopped.[29]

What is essential to our counterexample then, is that the behaviour of the mechanism gets fixed across close possible worlds. Anything else, including the intentions of its designers, is simply irrelevant. In fact we introduced the story of the cult into the example to ensure that the behaviour of its mechanism gets fixed across close possible worlds, but other stories could be told. Perhaps the cult intended to symbolise the cursed nature of the hour with a seemingly unnatural phenomenon. Indeed we could dispense with the cult entirely and stipulate that a bug in the programming of the microchip circuit of your clock causes it run perfectly reliably backwards from 5:00 pm to 4:00 pm during a particular hour.

4. Luck as Accident

We claimed that as in *Stopped Clock*, you do not know that it is 4:30 pm in *Backward Clock* because in both cases, your belief that it is 4:30 pm is luckily true. In both cases you were lucky to look at the position of the hands of your clock at the only instant during the hour when you nap at which they could have pointed

[27]Adams, Barker and Clarke, "Beat the Clock," 358.
[28]Adams, Barker and Clarke, "Beat the Clock," 357-358.
[29]Williams and Sinhababu, "The Backward Clock," 49.

to the correct time. Adams, Barker and Clarke announce that they "will take this to mean that the belief is accidentally true."[30] Then they say of *Backward Clock* that

> … there is luck/accident in the link from the actual time to the clock's displaying 4:30, and there is luck/accident in the link from the clock's displaying 4:30 to the subject's belief. Hence, the reason the subject lacks knowledge is that neither the clock nor his belief is tracking the truth. His belief is that it is 4:30, and it happens to be 4:30. But it is not the case that he believes it is 4:30 *because* it is 4:30 – his believing it to be 4:30 is not explained by the fact that it is 4:30.[31]

I do not see why we should prefer 'accidentally true belief' over 'luckily true belief'. The latter locution is more apposite because there is good and bad luck, and given that true beliefs are good, it is good luck in both *Stopped Clock* and *Backward Clock* that you end up with a true belief. Nonetheless let us talk of 'accident' as Adams, Barker and Clarke do. As we just saw in the last section, they take this as a synonym of 'coincidence,' in other words its being the case that two events or states of affairs happen to occur or obtain together, but without either causing the other. In both *Stopped Clock* and *Backward Clock* it is a coincidence that the time to which its hands point – 4:30 pm – is the time at which you look at them. Neither causes the other. But against Adams, Barker and Clarke, it is no coincidence that you acquire the belief that it is 4:30 pm when you look at its hands pointing to 4:30 pm. Your observation of the position of its hands, itself determined by their actual position, together with your understanding of how such positions represent time and your knowledge that your clock has always worked perfectly reliably, is what *makes you believe* that it is 4:30 pm.

Now they immediately conclude that

> So the method (or reason) that gives rise to the subject's belief is not a truth-tracking method (or reason), for it is too equivocal to yield knowledge.[32]

Let us postpone the question of what a truth-tracking method – as opposed to a truth-tracking belief – might be, until section 6. There remains their claim that your method of ascertaining the time you wake is 'equivocal.' What does this mean? In all three cases your method of ascertaining the time during the hour you nap is to look at your clock, or to put this more accurately, to observe the position of its hands. How is this 'equivocal' in *Backward Clock*? Their answer is that

[30]Adams, Barker and Clarke, "Beat the Clock," 357.
[31]Adams, Barker and Clarke, "Beat the Clock," 359.
[32]Adams, Barker and Clarke, "Beat the Clock," 359.

> … the clock's face emits false testimony for 59 minutes during that hour from 4:00 to 5:00. There is no way one could tell from this equivocal messenger which clock display (if any) expresses a truth rather than a falsity.[33]

To avoid the metaphor 'testimony,' this says that its hands point to the correct time only once during the hour that you nap. This is true. It is also true that you cannot tell, *just by observing the position of the hands*, when these point to the correct time. To know that, you would have to use an independent check of the accuracy of the clock, such as another clock that you know is accurate. But this is equally true of *Normal Clock*. So if *Backward Clock* is equivocal then so is *Normal Clock*, and if this is an impediment to knowledge then you can't know that it is 4:30 pm in *Normal Clock*. But you can.

In any case, even if Adams, Barker and Clarke have succeeded in showing that you do not know that it is 4:30 pm in *Backward Clock*, all the better for our counterexample against Nozick's analysis, since (1)-(4), and in particular (3), remain true.

5. How Dretske's Early Analysis Gets Clocked Out as Well

Early in their reply, Adams, Barker and Clarke say that

> … your method of forming your belief is the procedure you use to acquire your reason for believing, and your reason affords you knowledge only if it is sensitive to the truth – it wouldn't obtain unless your belief were true.[34]

This appears to espouse what we might call *Dretske's early analysis* of knowledge, that

> *S* knows that *p* just in case
>
> (1) *S* believes that *p* (without doubt, reservation or question) on the basis of *R*.
>
> (2) *R* would not be the case unless *p* were the case.
>
> (3) Either *S* knows that *R*, or *R* is some experiential state of *S*.[35]

Here *R* is a reason that *S* has for believing that *p*. We nowhere talked of a reason. We did however argue that *Backward Clock* shows that replacing Nozick's sensitivity condition (3) with various formulations of the safety condition on knowledge – roughly that *S*'s belief could not easily be false – still predicts

[33]Adams, Barker and Clarke, "Beat the Clock," 358.
[34]Adams, Barker and Clarke, "Beat the Clock," 354.
[35]Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49, 1 (1971): 12–13.

knowledge where there is ignorance.[36] There we argued that in using any of the clocks, if you know that it is 4:30 pm on a 'basis,' then this must be that the hands point to 4:30 pm. Let us call this a 'reason.' Now let us modify *Backward Clock* slightly. We may suppose that as you observe the position of its hands, you believe that it is 4:30 pm without doubt, reservation or question, because you know that your clock has always worked perfectly reliably. You base that belief upon your conjunctive reason that the hands point to 4:30 pm and your clock has always worked perfectly reliably. But this conjunction would not be true unless it were 4:30 pm, because the hands would not point to 4:30 pm unless it were 4:30 pm. This is because the circumstances in which you find yourself include those in which the clock runs perfectly reliably backwards from 5:00 pm to 4:00 pm. Finally, we may stipulate that you know the conjunction that the hands point to 4:30 pm and your clock has always worked perfectly reliably. (1)-(3) are all true, but you do not know that it is 4:30 pm any more than you know this in *Stopped Clock*. So Dretske's early analysis is also too weak, predicting knowledge where there is ignorance.[37]

6. What Are Truth-tracking Methods?

We have already noted that Adams, Barker and Clarke talk of truth-tracking methods, insisting that your method of ascertaining the time you wake is not truth-tracking. They also talk of sensitive methods.[38] However it is difficult to see how recourse to methods will block *Backward Clock*. The method you use to ascertain the time you wake is the same in all three case, namely to observe the

[36]Williams and Sinhababu, "The Backward Clock," 52–55.

[37]Tamar Lando, "Conclusive Reasons and Epistemic Luck," *Australasian Journal of Philosophy* 94, 2 (2015): 378–395, gives other counterexamples to this analysis. Dretske has a *later analysis*, that *K* knows that *s* is *F* just in case *K*'s true belief that *s* is *F* is caused (or causally sustained) by the information that *s* is *F* that is carried by a signal *r*, where *r* carries the information that *s* is *F* just in case the conditional probability of *s*'s being *F*, given *r* and *K*'s background knowledge is 1, but less than 1 given *K*'s background knowledge alone. See Fred Dretske, *Knowledge and the Flow of Information* (Cambridge, M.A.: MIT/Bradford, 1981), 86. *Normal Clock* seems to refute this. Your true belief that the time is 4:30 pm is supposed to be caused by the information that the time is 4:30 pm that is carried by the signal of the positions of the hands of your clock. But the conditional probability of the time being 4:30 pm, given these positions plus your background knowledge only that the clock has always worked perfectly reliably, is less than one. You cannot exclude the possibility that on this occasion the clock is not working reliably. Yet you know that it is 4:30 pm. Thus Dretske's later analysis is too strong, predicting ignorance where there is knowledge.

[38] For example, "This method is sensitive to the truth, for the clock wouldn't say what is says if the correct time weren't 4:30 p.m." (Adams, Barker and Clarke, "Beat the Clock," 355.)

position of the hands of your clock during the hour that you nap.[39] Since this method provides you with knowledge in *Normal Clock*, how can it deprive you of it in *Backward Clock*? Perhaps Adams, Barker and Clarke will reply that the method is sensitive in *Normal Clock* but not in *Backward Clock*. But they don't tell us what a sensitive method amounts to, at least not in a way that distinguishes it from a sensitive belief formed via a method. They say that

> Knowledge may be obtained, on tracking theories of knowledge, when p is true, one believes that p on the basis of a reason R or method M, and one's reason R (Dretske) or *method M* (Nozick) is *sensitive* to the truth – i.e. if p were not true one would not believe that p via R or M.[40]

This formulates a sensitive method as one that produces a sensitive belief, namely a belief that one wouldn't have were its content false. This doesn't help them, because the method you use in *Backward Clock*, namely observing the positions of its hands during the hour you nap, *does* produce a sensitive belief when you use it at 4:30 pm. Adams, Barker and Clarke need to make sensitive methods come apart from sensitive beliefs and then show that although your belief is sensitive, it is not produced by a sensitive method. Then they can go on to use this result to defend an analysis of knowledge that they seem to have in mind, in terms of sensitive, and presumably, truth-adherent methods. This *ABC analysis* (Adams, Barker and Clarke) would be as follows.

---

[39]Moreover this is how you would describe your method, which seems like a fair default way to decide what it really is (See John N. Williams, "Propositional Knowledge and Know-How," *Synthese* 165, 1 (2008): 122). We thought that this was clear in "Backward Clock" since we described the method identically in all three clock examples. In each case we started with "You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock…" It seems pretty clear from this that your method of ascertaining the time you wake is to observe, *during the period from 4:00 pm to 5:00 pm* (since that is the period during which you nap, not knowing when you will wake) the position of its hands. I am therefore surprised that Adams, Barker and Clarke countenance us as claiming that your method is 'looking at the clock at 4:30' (Adams, Barker and Clarke, "Beat the Clock," 360). Of course you wouldn't describe your method as this but as 'looking at the clock during the hour that I nap,' as they note (Adams, Barker and Clarke, "Beat the Clock," 360). They even claim that we described the method this way when Adams presented his reply in a talk in March 2016 (Adams, Barker and Clarke, "Beat the Clock," 360, note 14). We don't remember it that way. I suspect that they may have confused the claim they falsely attribute to us with our correct point that when you use your method of ascertaining what time you wake by observing the position of the hands *during the hour that you nap*, but *use it at 4:30 pm* then, in *Backward Clock*, you acquire a sensitive belief.

[40]Adams, Barker and Clarke, "Beat the Clock," 354, my italics.

> *S* knows that *p*, using method *M* of arriving at a belief whether *p*, just in case
>
> (1) *p*
>
> (2) *S* believes, using *M*, that *p*.
>
> (3) *M* is a sensitive method.
>
> (4) *M* is a truth-adherent method.

Since Adams, Barker and Clarke have no quarrel with the truth-adherence of your belief in *Backward Clock*, let us grant them (4) and concentrate on (3). What might a sensitive method be? Presumably it will have some connection with sensitive beliefs.[41]As just shown, they cannot say that a sensitive method is one that sometimes produces a sensitive belief. The alternative options are that it mostly or always produces sensitive beliefs. But now consider a fourth clock, one that combines *Normal Clock* with *Stopped Clock*. This is *Recently Stopped Clock*, as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand. Awaking at 4:55 pm, you see that its hands point to 4:55 pm. Accordingly, you form the belief that it is 4:55 pm. And it is indeed 4:55 pm because the clock has continued to work perfectly reliably until 4:50 pm, when a bug in the programming of its microchip circuit caused its hands to jump to 4:55 pm and then stop.

[41]Abandoning the connection looks unpromising. They cannot say that a sensitive method is *simply* one that always or mostly results in true beliefs. That would be to abandon truth-tracking altogether in favour of reliabilism, with its attendant difficulties, including the generality problem as raised by Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," *Philosophical Studies* 89, 1 (1998): 1–29. They can however claim that there is a *connection* between a sensitive method, whatever that will turn out to be, and a reliable one. In an exchange originating from Tristan Haze's apparent counterexamples to Nozick's analysis ("Two New Counterexamples to the Tracking Theory of Knowledge," *Logos & Episteme. An International Journal of Epistemology* VI, 3 (2015): 309–311), Haze reads Fred Adams and Murray Clarke ("Two Non-Counterexamples to Tracking Theories of Knowledge," *Logos & Episteme. An International Journal of Epistemology* VII, 1 (2016), 67–73) as saying that *S* knows that *p* via *M* just in case for all *q*, *S* believes that *q* via *M* just in case *q* (Haze, "Reply to Adams and Clarke," 225). I don't read them that way. This sees them as saying *inter alia*, that you know via a method only if that method allows you to *believe all truths*. Clearly that's far too strong. Adams and Clarke say that both Nozick's and Dretske's early analysis involve a method or a reason that is completely reliable (Adams and Clarke, "Two Non-Counterexamples," 71). So perhaps they think that a truth-tracking method, whatever that turns out to be, always produces true beliefs.

Most of the beliefs that you might form by observing the positions of the hands of your clock during the hour that you nap are sensitive. These are those that you would form during the period from 4:00 pm to 4:50 pm when it functions as *Normal Clock*.[42] So if a sensitive method is one that produces mostly sensitive beliefs, then the method that produces your belief that it is 4:55 pm is a sensitive method. Assuming that this method is also truth-adherent, the ABC analysis predicts that you know that it is 4:55 pm. But you don't, any more than you know that it is 4:30 pm in *Stopped Clock*. In both cases your belief is luckily true. You were lucky to look at your clock at 4:55 pm, at the only instant during the period from 4:50 pm to 5:00 pm at which its hands could have pointed to the correct time.

Now suppose that the clock's mechanism behaves in exactly the same way, but that waking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the true belief that it is 4:30 pm. Surely you know that it is 4:30 pm, because that is what you know in *Normal Clock* – which from 4:00 pm to 4:50 pm is essentially the same as your clock. But not all of the beliefs that you might form by observing the positions of the hands of your clock during the hour that you nap are sensitive. At 4.55 pm you might form the belief that it is 4.55 pm. That is an insensitive belief, because had you looked at your clock at 4:56 pm, then you would still believe that it is 4:55 pm. So if a sensitive method is one that always produces sensitive beliefs, then the method that produces your belief that it is 4:30 pm is an insensitive method. So (3) of the ABC analysis is false, with the result that it is now too strong, predicting ignorance where there is knowledge.

7. Concluding Remarks

Backward Clock shows that despite the arguments of Adams, Barker and Clarke, Nozick's analysis – as well as Dretske's early analysis – is too weak, predicting knowledge where there is ignorance. An analysis in terms of truth-tracking methods rather than truth-tracking beliefs is no remedy. One important conclusion that might be drawn from this discussion is that combining sensitivity conditions with truth-adherence conditions isn't enough to exclude epistemic luck.

[42] All of these beliefs are also true. This suggests that if a sensitive method is one that mostly produces sensitive beliefs, then a sensitive method mostly produces true beliefs. This however holds uninterestingly. A belief is sensitive just in case if *counterfactually false*, one wouldn't have it, so all sensitive beliefs are by definition true.

ARE GETTIER CASES MISLEADING?

Philip ATKINS

ABSTRACT: The orthodox view in contemporary epistemology is that Edmund Gettier refuted the JTB analysis of knowledge, according to which knowledge is justified true belief. In a recent paper Moti Mizrahi questions the orthodox view. According to Mizrahi, the cases that Gettier advanced against the JTB analysis are misleading. In this paper I defend the orthodox view.

KEYWORDS: analysis of knowledge, Gettier cases, semantic reference, speaker's reference

The orthodox view in contemporary epistemology is that Edmund Gettier refuted the JTB analysis of knowledge, according to which knowledge is justified true belief.[1] In a recent paper Moti Mizrahi questions the orthodox view.[2] According to Mizrahi, the cases that Gettier advanced against the JTB analysis are misleading. For, according to Mizrahi, they are cases of semantic failure (i.e., failure to refer to something) rather than epistemic failure (i.e., failure to know something). In this paper I defend the orthodox view.

I should mention that Mizrahi discusses several 'Gettier cases' besides the two that Gettier originated. Mizrahi discusses Roderick Chisholm's sheep case, Alvin Goldman's fake barn case, and Bertrand Russell's stopped clock case.[3] It is Mizrahi's opinion that all these cases are misleading. I disagree across the board, but for the sake of brevity I focus exclusively on Gettier's two cases. These are genuine counterexamples to the JTB analysis, or so I contend.

In Gettier's first case, Smith comes to have strong evidence for believing that Jones is the man who will get the job and that Jones has ten coins in his pocket. Smith makes a rudimentary logical inference and says the following:

(I) The man who will get the job has ten coins in his pocket.

---

[1] Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121-123.

[2] Moti Mizrahi, "Why Gettier Cases Are Misleading" *Logos & Episteme. An International Journal of Epistemology* VII, 1 (2016): 31-44.

[3] Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice Hall, 1966); Alvin Goldman, "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* 73 (1976): 771-791; Bertrand Russell, *Human Knowledge: Its Scope and Limits* (London: George Allen & Unwin, 1948).

It turns out that (I) is true, but not for the reasons that Smith thinks. For it turns out that Smith himself is the man who will get the job and that, unbeknownst to Smith, he also has ten coins in his pocket. Many have the strong intuition that Smith fails to know (I). Since Smith is justified in believing (I), we seem to have a counterexample to the JTB analysis.

But this case is misleading, according to Mizrahi, because Smith wishes to refer to the coins in Jones's pocket. Mizrahi invokes Saul Kripke's famous distinction between semantic reference and speaker's reference.[4] Roughly, the semantic referent of an expression is the thing designated by the expression according to the conventions of the language. The speaker's referent of an expression is the thing to which the speaker wishes to refer. Mizrahi argues that Gettier's first case is one where the semantic referent of 'coins' differs from the speaker's referent of 'coins.' The semantic referent is the set of coins in Smith's pocket, whereas the speaker's referent is the set of coins in Jones's pocket.[5] Since Smith fails to refer to the semantic referent of 'coins,' Gettier's first case is an instance of reference failure. This makes the case misleading, according to Mizrahi, since "we may be confusing the fact that Smith fails to refer to what actually fulfills the conditions for being the semantic referent of 'coins,' which is a semantic fact about the case, with an epistemic fact, namely that Smith doesn't know that (I) is the case."[6] Mizrahi concludes that our intuition about Gettier's first case should not be assigned much evidential weight.

Central to Mizrahi's argument is the possibility that we are confusing a certain kind of semantic failure with a certain kind of epistemic failure. This is supposed to explain our intuition about Gettier's first case. But if this is the correct explanation of our intuition, then the intuition should be absent when there is no such semantic failure. Unfortunately for Mizrahi, it is easy to revise Gettier's first case so that there is no such semantic failure. Suppose that Smith has strong evidence for believing that Jones is the man who will get the job and that Jones is handsome. We can suppose that Smith is justified in believing that Jones is handsome based on seeing Jones in person. Smith makes a rudimentary logical inference and says the following:

(I*) The man who will get the job is handsome.

---

[4] Saul Kripke, "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy* 2 (1977): 255-276.

[5] Actually, it is far from clear that the semantic referent of 'coins' should be identified the specific set of coins in Smith's pocket, but this is Mizrahi's assertion, which I am willing to accept for present purposes.

[6] Mizrahi, "Why Gettier Cases," 35.

It turns out that (I*) is true, but not for the reasons that Smith thinks. For it turns out that Smith is the man who will get the job and that, unbeknownst to Smith, he is also handsome. I cannot speak for everyone, but I have the strong intuition that Smith fails to know (I*). Since Smith is justified in believing (I*), we seem to have a counterexample to the JTB analysis.

As far as I can tell, there is no semantic failure when Smith uses the predicate 'is handsome.' Of course, theorists have various different opinions regarding the semantics of predicates. But, according to standard accounts, the semantic referent of a monadic predicate is a certain property, set, or function. To simplify matters, let us assume that the semantic referent of 'is handsome' is the property *handsomeness*. There is no special reason to insist that Smith, when using this predicate, fails to designate this property. He is using the predicate in the same way that he typically uses the predicate, after all. Mizrahi might argue that Smith is referring to *Jones's handsomeness*, rather than the general property *handsomeness*, and therefore there is semantic failure. But this argument would have very little plausibility. There are no independent grounds for making this argument, except that Smith intends to apply the predicate 'is handsome' to Jones. From this fact alone we should not conclude that there is semantic failure, unless we are prepared to conclude that many (most?) ordinary uses of the predicate are instances of semantic failure. When I say 'The president of the United States is handsome,' I intend to apply the predicate to Barack Obama. When I say 'Ryan Gosling is handsome,' I intend to apply the predicate to Ryan Gosling. When I say 'That guy is handsome,' I intend to apply the predicate to that guy. Even though I have specific men in mind when I use the predicate, it would be inappropriate to insist that I have failed to designate the semantic referent of the predicate.

In order for my case to parallel Gettier's original case, I have stipulated that Smith does not know that the predicate 'is handsome' applies to himself.[7] This might strike some readers as artificial, but it is not unrealistic to suppose that Smith is humble and therefore unaware of his own attractiveness. Of course, we can set up the case using different predicates. It seems to me that any monadic predicate would suffice, so long as Smith is justified in believing that the predicate applies to Jones, and Smith is unaware that the predicate also applies to himself. For example, we can set up the case using the predicate 'is wealthy.' For we can suppose that Smith is justified in believing that Jones is wealthy, but that Smith is

[7] Personally, I do not believe that this stipulation is necessary to refute the JTB analysis. Even if the case is set up so that Smith knows that he is handsome, I have the intuition that Smith does not know (I*). Similarly, even if Gettier's original case is set up so that Smith knows that he has ten coins in his pocket, I have the intuition that Smith does not know (I).

unaware that he himself is wealthy (unbeknownst to Smith, he has recently inherited a fortune from a deceased relative). We can set up the case using the predicate 'is Canadian.' For we can suppose that Smith is justified in believing that Jones is Canadian, but that Smith is unaware that he himself is Canadian (Smith is suffering from selective amnesia). None of these cases are plausible examples of semantic failure. All of these cases refute the JTB analysis of knowledge.

But perhaps I am being too quick. Even though there is no semantic failure when Smith uses the relevant predicate, it is somewhat plausible that there is semantic failure when Smith uses the definite description 'the man who will get the job.' The speaker's referent is Jones, whereas the semantic referent is Smith himself. This is not the kind of semantic failure emphasized by Mizrahi, but others have argued that it undermines Gettier's first case.[8] Does the possibility of this kind of semantic failure show that our intuition about Gettier's first case should not be assigned much evidential weight? I think not. If this kind of semantic failure is the correct explanation of our intuition, then the intuition should be absent when there is no such semantic failure. But again it is easy to revise Gettier's first case so that there is no such semantic failure. Suppose again that Smith has strong evidence for believing that Jones is the man who will get the job and that Jones is handsome. Smith performs an existential generalization and says the following:

(I**) There is someone who is both getting a job and handsome.

It turns out that (I**) is true, but not for the reasons that Smith thinks. For it turns out that (I**) is made true by Smith himself. Even though Smith is justified in believing (I**), and even though (I**) is true, I have the strong intuition that Smith fails to know (I**). This intuition cannot be explained by the kind of semantic failure discussed above, since no such semantic failure is present. We are dealing here with an existential generalization. There is no definite description whose semantic referent is Smith and whose speaker's referent is Jones. There is no definite description whatsoever.

[8] Adrian Heathcote, "Truthmaking and the Gettier Problem," in *Aspects of Knowing: Epistemological Essays*, ed. Stephen Hetherington (Amsterdam: Elsevier, 2006), 151-168. Also relevant is Christoph Schmidt-Petri, "Is Gettier's First Example Flawed?" in *Knowledge and Belief*, ed. W. Löffler and P. Weingartner (ALWS, 2003), 317-319. Note that Schmidt-Petri relies on Keith Donnellan's distinction between the *referential* use of a description and the *attributive* use of a description, rather than Kripke's more general distinction between speaker's reference and semantic reference. See Keith Donnellan "Reference and Definite Descriptions," *The Philosophical Review* (1966): 281-304.

Of course, there is some sense in which Smith has Jones in mind when inferring (I**), but this point seems irrelevant. The fact of the matter is that (I**) is an existential generalization, whose content does not include Jones (or anyone else). There is nothing to prevent Smith from believing (I**). Since this belief is justified and true, the relevant question is whether this belief counts as knowledge. Unless we insist that Smith does indeed know (I**), we must conclude with Gettier that the JTB analysis is false.[9]

I turn finally to Gettier's second case. In this case Smith has strong evidence for believing that Jones owns a Ford. His evidence is that "Jones has at all times in the past owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford."[10] Smith makes a rudimentary logical inference and says the following:

(h) Either Jones owns a Ford or Brown is in Barcelona

In fact, Smith does not know Brown's location. Still, Smith is justified in believing (h), since Smith inferred it from something else that he is justified in believing. It turns out that (h) it true, but not for the reasons that Smith thinks. The twist in this case is that Jones does not own a Ford, but, by sheer coincidence, Brown is in Barcelona. Many have the strong intuition that Smith fails to know (h). Since Smith is justified in believing (h), we seem to have a counterexample to the JTB analysis.

This case is misleading, according to Mizrahi, because Smith wishes to refer to the person who has always owned a Ford, who has just offered him a ride while driving a Ford, and so on. This is the speaker's referent of 'Jones.' Mizrahi argues that the speaker's referent of 'Jones' must differ from the semantic referent of 'Jones.' Unfortunately for Mizrahi, his argument betrays a serious misunderstanding of Gettier's second case. Mizrahi sets up the case so that Smith comes to have evidence for believing

(f) Jones owns a Ford

[9] Someone might insist that Smith knows (I**) on independent grounds: there are so many handsome men in the world, it is reasonable to think that at least one of them is getting a job. But we can easily avoid this complication by replacing 'is handsome' with a predicate that applies to fewer people. Then Smith's only justification for believing (I**) would have to do with Jones. Even then someone might insist that Smith knows (I**). I suspect that Christoph Schmidt-Petri would insist that Smith knows (I**), though I am not certain (see "Is Gettier's First Example Flawed?"). Since Schmidt-Petri's remarks on Gettier are complicated, and since I am concerned specifically with Mizrahi's argument, I must set the matter aside. Readers can decide for themselves whether it is plausible to insist that Smith knows (I**).

[10] Gettier, "Is Justified True Belief Knowledge," 122.

and then infers

(g) Either Jones owns a Ford or Brown is in Boston

and then, from (g), infers (h). According to Mizrahi, it is important that this case involves two separate inferences, the first from (f) to (g), and the second from (g) to (h). He argues that "the speaker's reference of 'Jones' is the person who offered Smith a ride, has always owned a Ford, etc., whereas the semantic referent of Jones, i.e., the referent of 'Jones' that makes (g) true, cannot be that person, since Jones does not own a Ford, by stipulation."[11] In the first place, it is not clear why Mizrahi thinks that (g) must be true. But the more salient problem is that Gettier's case does not involve an inference from (g) to (h). Gettier explicitly presents the case so that (h) is inferred directly from (f), by disjunction introduction. The source of Mizrahi's confusion seems to be that Smith also infers (g) from (f). But this is an inessential feature of the case. Gettier introduces (g) only to emphasize the fact that Smith is selecting cities at random. In fact, as Gettier presented the case, Smith also infers the following from (f):

(i) Either Jones owns a Ford or Brown is in Brest-Litovsk.

Again we are not supposed to think that (i) is inferred from (g) or from (h). Instead we are supposed to think that (g), (h), and (i) are each inferred from (f). Smith believes each disjunction, and he is justified in believing each disjunction, since they are each inferred from something else that he is justified in believing. But only (h) happens to be true (unbeknownst to Smith!).

So, contrary to Mizrahi's presentation of the case, (g) is not true, and (h) is not inferred from (g). Furthermore, there are not two separate men, one of whom is the speaker's referent of 'Jones' and one of whom is the semantic referent of 'Jones.' There is only one man, the subject of (f). The case is such that Smith is justified in believing (f), even though (f) is false. We can assume that (f) is false because Jones has lost ownership of his old Ford and "is at present driving a rented car."[12] Nonetheless, Smith has correctly inferred (h) from (f), and (h) is made true by the fact that Brown is in Barcelona. There is no semantic failure in this case. When Smith uses 'Jones,' he is successfully and consistently referring to a single man, the semantic referent of 'Jones.' Mizrahi has provided no reason to suspect that our intuitions about this case are inaccurate, or that this case is ambiguous in any important respect.[13]

[11] Mizrahi, "Why Gettier Cases," 36.

[12] Gettier, "Is Justified True Belief Knowledge," 123.

[13] I am indebted to Rebecca Pluckhorn and Matt Griffin for helpful discussion.

ASSERTION: JUST ONE WAY TO TAKE IT BACK

Mona SIMION

ABSTRACT: According to Jonathan Kvanvig, the practice of taking back one's assertion when finding out that one has been mistaken or gettiered fails to speak in favour of a knowledge norm of assertion. To support this claim, he introduces a distinction between taking back the content of the assertion, and taking back the speech act itself. This paper argues that Kvanvig's distinction does not successfully face close speech-act-theoretic scrutiny. Furthermore, I offer an alternative diagnosis of the target cases sourced in the normativity of action.

KEYWORDS: assertion, excuse, knowledge norm, Kvanvig, taking back

1. Introduction

One must: assert that p only if one knows that p. Or at least that's what a very popular view on the epistemic normativity of assertion stipulates. This has become known in the literature as the Knowledge Norm of Assertion (KNA).[1] In spite its popularity, KNA is taken by some to be too strong a requirement. Jonathan Kvanvig, for instance, defends a weaker, justified belief norm on assertion (henceforth, JNA), where the relevant epistemic standing is knowledge-level justification.[2] It is argued that KNA, as opposed to JNA, has a hard time explaining cases in which assertions on some lesser epistemic standings do not render the speakers subject to criticism. Assertions on false belief that the speaker mistakes for knowledge and assertions on gettiered belief are cases in point.

Defenders of KNA have mostly employed one version or another of what has become known as the 'excuse manoeuvre.' Williamson, for instance, argues that speakers asserting on what they mistakenly take to be knowledge, although in

[1] This paper is only concerned with the necessity direction of the knowledge norm of assertion. The *locus classicus* for the defence of KNA is Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).

[2] E.g. Jonathan Kvanvig, "Assertion, Knowledge and Lotteries," in *Williamson on Knowledge*, ed. Patrick Grenough and Duncan Pritchard (Oxford: Oxford University Press, 2009), 140-160 and Jonathan Kvanvig, "Norms of Assertion," in *Assertion: New Philosophical Essays*, ed. Jessica Brown and Herman Cappelen (Oxford: Oxford University Press, 2011), 233-250.

© LOGOS & EPISTEME, VII, 3 (2016): 385-391

breach of the norm, have a good excuse for making an impermissible assertion.[3] One way to see this is by noticing that, as soon as they discover they have been mistaken or gettiered, speakers will typically take back their assertions.

In defence of JNA, Kvanvig distinguishes between two ways of taking back an assertion: by taking back the speech act itself, in cases in which the speaker lacks proper justification for his assertion, or by only taking back the content of the speech act – in cases of false or gettiered justified beliefs. The norm of assertion, Kvanvig argues, is a norm governing a type of human activity. Therefore, only when the act itself is taken back should we consider the norm to have been broken.

This paper is a rejoinder on behalf of KNA. It is argued that Kvanvig's distinction between two ways of taking back does not successfully face close speech-act-theoretic scrutiny. To show this, I will first introduce the target cases (section 2); further on, I briefly outline Kvanvig's 'taking back' argument and show why it fails (section 3). In section 4, I will offer a diagnosis of Kvanvig's cases sourced in the normativity of action, which will turn out to be perfectly compatible with KNA. In the last section I conclude (5).

2. Assertions from Belief that Falls Short of Knowledge

Consider the following two cases:

> (i) Assertion on justified false belief:
>
> FAKE SNOW: […] it is winter, and it looks exactly as it would if there were snow outside, but in fact that white stuff is not snow but foam put there by a film crew of whose existence I have no idea. I do not know that there is snow outside, because there is no snow outside, but it is quite reasonable for me to believe not just that there is snow outside but that I know that there is; for me, it is to all appearances a banal case of perceptual knowledge. Surely it is then reasonable for me to assert that there is snow outside.[4]

And

> (ii) Assertion on justified true belief that falls short of knowledge:
>
> FAKE BARNS: […] suppose that Wendy correctly sees the only real barn that, unbeknownst to her, is completely surrounded by barn facades and asserts to me "There was a barn in the field we just passed" on this basis.[5]

[3] Williamson, *Knowledge* and "Replies to Critics," in *Williamson on Knowledge*, ed. Patrick Grenough and Duncan Pritchard (Oxford: Oxford University Press, 2009), 279-385.
[4] Williamson, *Knowledge*, 257.
[5] Jennifer Lackey, "Norms of Assertion," *Nous 41* (2008): 544.

In both the cases above, speakers assert from what they mistakenly take to be knowledge. And, intuitively, they can hardly be subject to blame. Also, notice that no further normative constraints seem to be active in these cases, so as to maybe override the epistemic requirement.

In defence of KNA, Williamson argues that, in the cases above, although the speaker has a good excuse for having broken KNA, he is still in breach of the norm. According to him, it would seem natural for someone who had strong reasons to think what he asserted was true, to apologize when finding out it was actually false. Here is Williamson:

> Misrecognizing someone, I may say: "That's Sasha – no, sorry, it's not – it's just someone who looks very like him." […] Nor is it strange for a newspaper to apologize to its readers for an error in a previous edition, nor for the author of a book to apologize in the preface for any remaining errors, even though every effort has been made to ensure that the contents are correct.[6]

Equally, it would not seem very odd if Wendy, after you point out to her that she's in Fake Barn County, were to say something along the lines of: "Sorry, I didn't know that."

3. Two Ways of Taking Back

Notice, however, that Williamson's defence fails to establish that excuses, while not odd, are really *necessary* in these cases. However, one thing is clear: after finding out that one was mistaken or gettiered, one should at least not stand by the commitments implied by one's assertion anymore. Thus, rather than presenting excuses as such, an appropriate reaction would go along the lines of "Oh, I take that back. I was not aware of there being a film crew producing fake snow outside," or "Oh, I take that back, I had no idea we were in Fake Barn County."

In support of JNA, however, Jonathan Kvanvig distinguishes two types of attitude a speaker may have in response to her assertions being corrected. Kvanvig argues that "in some cases of correction, we take back the content of our speech act, and in other cases we apologize for, and regret, the very act itself." For example, if we assert p and then are shown that p is false, we take back the content of our speech act, but we needn't apologize for or regret the very act itself. "In fact, were [we] to apologize, the natural response would be dismissive: Give it a rest, nobody's always right..." According to Kvanvig, the same distinction plays out with gettiered assertions too. Thus, presumably, in the case of Wendy above, if

[6]Williamson, "Replies," 345.

after she asserts "There's a barn in the field," and I point out to her that she can't possibly know that, as we are in Fake Barn County, she would just take back *what* she said, not to apologize for having said it.

Kvanvig argues that things are different when you don't have justification for what you say, even if, by some bizarre twist, you turn out to be right. In support of this, he offers the case of Billy Bob, a Texas Democrat, who, based on a headline on a tabloid, asserts to his friend Sue: "George Bush is a communist!" When Sue points out to him that he should not trust tabloids, Billy Bob apologizes: "You're right, I shouldn't have believed that paper and I shouldn't have said what I did. I take it back."

According to Kvanvig, in this situation, apologizing and taking back the speech act itself is the right thing to do. He argues that norms of assertion are norms governing a certain type of human activity, and thus relate to the speech act itself rather than the content of such an act. As such, only when the speech act itself is at fault, do we have reason to think that some norm of assertion is broken; when only the content of the assertion needs to be taken back, the assertion itself is not at fault.[7]

Here is, however, some reason to doubt that Kvanvig's distinction works; speech act literature[8] distinguishes between the content of a speech act and the illocutionary force by which the content is being put forward. One can perform various speech acts upon p: one can ask whether p, promise that p, threaten that p, etc. In the case of assertion, by uttering p the speaker presents p as true.

Given this, a *proposition* is itself communicatively inert; that is to say that to actually perform a speech act, one has to put forth a proposition with an *illocutionary force,* such as assertion, promise, command, etc.

But if the propositional content is inert in isolation, it is less clear how Kvanvig envisages one being able to take it back in isolation. To see this, notice that assertion, as opposed to other types of actions – say, having vacationed in Hawaii – can be 'taken back.' Not in the sense that one can change the past as to not have had asserted in the first place, of course. Rather, taking back an assertion that p refers to no longer standing behind the commitments implied by having asserted that p. Now, p itself, in isolation, does not imply any commitments whatsoever. That is, depending on which illocutionary force we will act upon it with, different commitments will follow. If I promise that p, for instance, I

[7]Kvanvig, "Assertion, Knowledge and Lotteries," 148.

[8]See, e.g. Mitchell Green, "Speech Acts," *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/speech-acts/>. Accessed January 2016.

commit myself to a future course of action; if I assert that p, I commit myself to, at least, it being the case that p.

If that is the case, it becomes clear that in order to take an assertion back, that is, to be released from the commitments implied by it, it has to be the case that I take back everything, force and content. I cannot only take back the content p, because p in isolation does not commit me to anything, inasmuch as I do not present it as true, or command p, or promise p, etc. Also, I cannot only take the action back either, because presenting nothing as true, or promising nothing also fails to imply any commitments on my part.

So the only way in which one can take an assertion back is by not standing behind the commitments implied by the whole compound: having presented p as true.

4. Diagnosis

Something seems, indeed, intuitively different between the two cases presented by Kvanvig, though. To see what it is, let us start by clearing the normative air a bit.

According to a fairly uncontroversial view in the normativity literature[9] that has been with us since Aristotle, one is an apt candidate for blame for violating a norm only if the agent is aware of what it is she is doing or bringing about.[10] As such, one may reasonably do something impermissible because one reasonably but falsely believes it to be permissible. If your car's (well maintained) speedometer has unluckily just broken, you might break the norms of safe driving due to its misreadings, and still be blameless for doing so. For all you know, your act is proper according to the norm, even though, in fact, this is not the case. Similarly, if you fail to keep your promise to meet your friend Ted for lunch because your (otherwise highly reliable) secretary misinforms you about the time at which you're supposed to meet him, you're blameless for not showing up. However, your having broken your promise remains an improper act according to the norms of social commitment.

---

[9] See e.g. Ishtiyaque Haji, *Moral Appraisability* (Oxford: Oxford University Press, 1998) and Michael Zimmerman, "Moral Responsibility and Ignorance," *Ethics 107* (1997): 410-426. People working in this field disagree whether a belief or a knowledge condition is appropriate for blameworthiness. Although not much in this paper hinges on this, I here go with the stronger view – supporting the belief condition – both because I find it more plausible, and in order to stay on the safe side by attributing blameworthiness more generously.

[10] Aristotle, *The Nicomachean Ethics* (Terrence Irwin, transl., Indianapolis: Hackett Publishing Co, 1985), 1110a-1111b4.

With regard to this though, some qualifications are needed. The literature[11] distinguishes between direct and indirect blameworthiness for performing an action. One is indirectly blameworthy for something x, if and only if one is blameworthy for it by way of being blameworthy for something else, y, of which x is the consequence.

One could be indirectly blameworthy for performing an action out of ignorance, by being directly blameworthy for being ignorant. Notice, though, that in both the above cases, although the agent ends up with a false belief that his actions are in accordance with the relevant norms, this seems to happen through no fault of his own. That is, he seems to have conformed to his epistemic duties: coming to believe that you are driving at a certain speed via looking at your car's well maintained speedometer is a quite reliable way to go about it, as is asking your secretary about your schedule for the day. Surely, if our agent were to be speeding due to his trusting his three years old son's readings of the speedometer, we would tend to find him blameworthy for his breaking the traffic norms.[12] Thus, let us formulate the principle governing the relationship between awareness of breach of the norm and blameworthiness as follows:

> **Blame-Awareness**: An agent is blameless for performing an all-things-considered improper act if she conformed to her epistemic duties and she had good reasons to believe she was respecting the norm.

Let us now, in the light of this, go back to the cases put forth by Kvanvig. First, by *Blame-Awareness*, the speakers in FAKE SNOW and FAKE BARNS, asserting on justified belief, are epistemically blameless, both directly and indirectly. They both assert from what they mistakenly take to be knowledge, and they seem to have conformed to their epistemic duties in forming the respective beliefs. After all, perception is a pretty reliable way to go about forming beliefs. In contrast, notice that Billy Bob's belief formation process, as Sue rightly points out, does not stand very tall when it comes to reliability. So, indeed, Billy Bob is indirectly blameworthy, as he failed to conform to his epistemic duties before proceeding, which led to him being in breach of the norm.

[11] E.g. Zimmerman, "Moral Responsibility."

[12] It might also be that your belief is unjustified yet blameless – say because you have been brainwashed into believing your 3-year-old son on this. This case, however, concerns a control condition on blameworthiness that falls outside the scope of this paper. I discuss it more in detail in Mona Simion, "Knowledge, Rational Credibility and Assertion: The Scoreboard," in *Epistemic Reasons, Epistemic Norms and Epistemic Goals,* ed. Martin Grajner and Pedro Schmechtig (DeGruyter Berlin/Boston, forthcoming).

5. Conclusion

I have argued that Kvanvig's distinction between two ways of taking back a speech act does not successfully face close speech-act theoretic scrutiny. In order to be released from the commitments implied by a speech act, one has to take back both content and illocutionary force; one without the other will not imply any commitments to begin with. Also, I have put forth a KNA-friendly explanation of Kvanvig's target cases sourced in the normativity of action in general, so as to stay off suspicions of ad-hocness.

NOTES ON THE CONTRIBUTORS

Fred Adams is Professor of Linguistics and Cognitive Science and Professor of Philosophy at the University of Delaware. He has published well over 100 articles and chapters on cognitive science, epistemology, philosophy of language, philosophy of action, and related topics. He is also co-author of *The Bounds of Cognition* (Wiley-Blackwell, 2008, with Ken Aizawa) and of *An Introduction to the Philosophy of Psychology* (Cambridge University Press, 2015, with Daniel Weiskopf. Contact: fa@udel.edu.

Philip Atkins is currently a Visiting Assistant Professor of Philosophy at Temple University. He received his doctorate in 2013 from the University of California, Santa Barbara. His research interests include the philosophy of language, epistemology, and metaphysics. Some of his recent publications include "A Russellian Account of Suspended Judgment" (*Synthese*) and "A Pragmatic Solution to Ostertag's Puzzle" (*Philosophical Studies*). He is also the coauthor of "Essential Versus Accidental Properties" (*Stanford Encyclopedia of Philosophy,* with Teresa Robertson). Contact: philip.atkins@temple.edu.

John A. Barker is Professor Emeritus of Philosophy at Southern Illinois University Edwardsville. He is the author of numerous articles in the areas of epistemology, philosophy of language, logic, philosophy of mind, and philosophy of science. He is also the author of a book, *A Formal Analysis of Conditionals* (Southern Illinois University Monographs, 1969) and of a software application, *ProtoThinker: A Model of the Mind* (Wadsworth, 1998 and ProtoThinker.com, 2016). Contact: johnabarker1@cs.com.

Constantin C. Brîncuş is a PhD candidate at the Faculty of Philosophy, University of Bucharest. His main research interests include modern logic, philosophical logic, philosophy of logic, philosophy of language and epistemology. His most recent articles include "Inferring to the Best Explanation of First-Order Logical Truths" (forthcoming in *The Romanian Journal of Analytic Philosophy*, 2016), "The Epistemic Significance of Valid Inference – A Model-Theoretic Approach," in *Meaning and Truth*, ed. Sorin Costreie and Mircea Dumitru (Bucharest: Pro Universitaria Publishing House, 2015). "A Carnapian Approach to Counterexamples to Modus Ponens," *Romanian Journal of Analytic Philosophy* (2013, with Iulian D. Toader). Contact: c.brincus@yahoo.com.

Lajos Brons is an adjunct professor of philosophy at Lakeland University, Japan

Campus, and a researcher at the philosophy department of Nihon University, both in Tokyo, Japan. His research is divided over two broad areas: one is in the overlap of ethics and social/political philosophy; the other is in the intersection of philosophy of language, metaphysics, and epistemology. His recent publications include "Putnam and Davidson on Coherence, Truth, and Justification" (*The Science of Mind*, 2016), "Othering, an Analysis" (*Transcience*, 2015), "Wang Chong, Truth, and Quasi-Pluralism" (*Comparative Philosophy*, 2015), and "Meaning and Reality: a Cross-Traditional Encounter" (Mou & Tieszen, *Constructive Engagement of Analytic and Continental Approaches in Philosophy*, 2013). Contact: mail@lajosbrons.net.

Murray Clarke is Professor in the Department of Philosophy at Concordia University in Montreal, Quebec. His research interests span contemporary epistemology, cognitive science and philosophy of science. He is the author of *Reconstructing Reason and Representation* (MIT, 2004). He has published numerous articles including "Resurrecting the Tracking Theories" (*Australasian Journal of Philosophy*, 2005), "Concepts, Intuitions and Epistemic Norms" (*Logos & Episteme. An International Journal of Epistemology*, 2010), and "Critical Notice: Jose Zalabardo's Scepticism and Reliable Belief" (*Teorema*, 2014). Contact: Murray.Clarke@concordia.ca.

Arnold Cusmariu is a philosopher and a working sculptor. He has published articles on topics in metaphysics, epistemology, the philosophy of mathematics, the philosophy of mind and aesthetics – available at www.Academia.edu. His World Trade Center Memorial Competition entry can be viewed at http://www.wtcsitememorial.org/ent/entI=446704.html. Also a screenwriter, his scripts *Muybridge*, *Fancy Free*, and *Light Becomes Her* are available at www.zoetrope.com. Contact: bravo323@gmail.com.

Mona Simion is a Doctoral Researcher at the Arché Research Centre, University of St. Andrews and the KU Leuven Epistemology Group. Her research focuses mainly on epistemology and ethics, but it also touches on issues in the philosophy of language. She has worked on topics like epistemic norms, the value of knowledge, perceptual and testimonial entitlement, epistemic contextualism, pragmatic encroachment, the transitivity of value ordering, distributive justice, meta-decision theory, criticism and blame. She is currently developing a function-first integrated account of epistemic norms for practical reasoning, assertion and belief. On this account, the epistemic norms at stake are taken to drop out of the respective epistemic functions. Contact: mis2@st-andrews.ac.uk.

Pierre Uzan is a member of SPHERE laboratory of History and Philosophy of Science, University Paris-Diderot/CNRS. His main areas of research are in Philosophy of Quantum Mechanics, Quantum Logic and Philosophy of Mind. He has suggested an explanation of the nature of psychophysical correlations in terms of (generalized) entanglement, without appealing to a concept of causality and without identifying mind and brain. His recent publications include *Conscience et physique quantique* (Vrin, Paris) and several articles bearing on the quantum-like modelling of the phychosomatic unity of the individual, like "On the nature of psychophysical correlations" (*Mind and Matter*, 2014). Contact: pierre.uzan@paris7.jussieu.fr.

Raphael van Riel is associate professor at the University of Duisburg-Essen (Germany) and principal investigator in the research group „A Study in Explanatory Power." His research mainly focuses on theories of explanation. He has published one book (*The Concept of Reduction*, Springer, 2014), he has co-authored the entry "Scientific Reduction" in the Stanford Encyclopedia of Philosophy (together with Robert N. Van Gulick), and he has published numerous papers and chapters on various topics, mainly in theoretical philosophy. Contact: Raphael.vanRiel@uni-due.de.

Kevin Wallbridge is a PhD candidate at the University of Edinburgh. His primary interests are in epistemology, although his research also touches upon issues in the philosophy of mind, metaphysics, and language. His PhD thesis, *Counterfactual Epistemology: in Defence of Sensitivity,* defends the claim that knowledge requires sensitive belief and examines how this claim interacts with anti-luck epistemology, virtue epistemology, hinge-epistemology, and contextualism. Contact: k.j.wallbridge@gmail.com.

John N. Williams is an associate professor in philosophy in the School of Social Sciences at Singapore Management University. His research interests include paradoxes, epistemology and applied ethics. He has published extensively on Moore's Paradox. His most recent articles include: "Still a New Problem for Defeasibility: A Rejoinder to Borges" (*Logos & Episteme. An International Journal of Epistemology*, 2016), "The Backward Clock, Truth-Tracking, and Safety" (with Neil Sinhababu, *Journal of Philosophy*, 2015), "Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge" (*Analysis*, 2015), "Moore's Paradox in Speech: A Critical Survey" (*Philosophy Compass*, 2015), and "Moore's Paradox in Thought: A Critical Survey" (*Philosophy Compass*, 2015). Contact: johnwilliams@smu.edu.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears in printed and on-line version at the end of March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal will publish articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal intends to promote all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

NOTES TO CONTRIBUTORS

1. Accepted Papers

The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

2. Submission Address

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*rtf) or Microsoft Word document (*docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author`s affiliation and professional address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it. The review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers. The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are

received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address:

Eugen Huzum

Institutul de Cercetări Economice şi Sociale „Gh. Zane”

Academia Română, Filiala Iaşi

Str. Teodor Codrescu, Nr. 2, 700481

Iaşi, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Property & Royalties

Articles accepted for publication will become the property of Logos & Episteme and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay

royalties. Authors of accepted manuscripts will receive free of charge one copy of the issue containing their papers.

14. Permissions

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.