Volume XII ◇ Issue 4

2021

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

http://logos-and-episteme.acadiasi.ro/

https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme

ISSN 2069-053

TABLE OF CONTENTS

RESEARCH ARTICLES

REDUCTION, INTUITION, AND COGNITIVE EFFORT IN SCIENTIFIC LANGUAGE

Miguel LÓPEZ-ASTORGA

ABSTRACT: In his search for a better scientific language, Carnap offered a number of definitions, ideas, and arguments. This paper is devoted to one of his definitions in this regard. In particular, it addresses a definition providing rules to add new properties to the descriptions of objects or beings by taking into account other properties of those very objects or beings that are already known. The main point that this paper tries to make is that, if a current cognitive theory such as the theory of mental models is assumed, it can be said that those rules are easy to use by scientists and philosophers of science. This is because, following the essential theses of this last theory, the rules do not demand excessive cognitive effort to be applied. On the contrary, they are simple rules that make researchers' work harder in no way.

KEYWORDS: Rudolf Carnap, dual-process theory, predicates, scientific language, theory of mental models

Introduction

It is well known that Rudolf Carnap tried to build a language for science.[1] Thus, he gave different definitions and proposals. This paper will be focused on one of such definitions.[2] That definition includes four formulae and three rules to link new predicates to elements from other relations between predicates already provided. Thereby, the idea is to express, by means of logical formulae, new relations between predicates in order to make it explicit that, if an object or being has one predicate, that object or being must also have another one. In this way, the rules allow introducing gradually new properties from just a few primitive properties in the language. The formulae are the following:

(1) IF Q_1 THEN (IF Q_2 THEN Q_3)

[1] E.g., Rudolf Carnap, "Testability and meaning," *Philosophy of Science* 3, 4 (1936): 419-471.

[2] Carnap, "Testability and meaning," 442-443; Definition 10.

With other symbols, (1) is (R) in Carnap's account.[3] 'IF…THEN…' stands for logical conditional relation, and, as (R) in Carnap's account,[4] (1) is universally quantified and the universal quantifier is omitted only for shorten. Accordingly, what (1) means is that, for every object or being, if that object or being has property Q_1, then if that object or being also has property Q_2, that object or being has property Q_3 too.

(2) IF Q_1 THEN (IF Q_2 THEN Q_3)

(3) IF Q_4 THEN (IF Q_5 THEN not-Q_3)

There is no doubt that (2) is identical to (1). However, Carnap presents (2) and (3) as a pair separated from (1). That is because, as shown below, in his framework, (1) is used to give a rule by itself, and (2) and (3) are taken together to offer another different rule. Thereby, with other symbols, (2) and (3) are (R_1) and (R_2) in Carnap's account,[5] and they are, as (1), universally quantified sentences.

(4) IF Q_1 THEN (Q_3 IFF Q_2)

Clearly, '…IFF…' is an abbreviation for '…IF AND ONLY IF…', it represents logical biconditional relation, and (4), which is (R_b) in Carnap's account,[6] is a universally quantified sentence as well.

These formulae are interesting because Carnap[7] indicates limitations for them, and those limitations can reveal several points. For example, an analysis of them can show that (1), (2), (3), and (4) may not be so demanding for scientists or philosophers of science. At least, if a theory such as the theory of mental models[8] is assumed as the approach that describes the actual manner human beings think, reason, and make inferences.

This is what will be shown then. First, what the three rules and four formulae are and what their restrictions are will be explained in more detail. Second, relevant aspects of the theory of mental models for the analysis of (1), (2), (3), and (4) will be commented on. Lastly, the reasons why, under the theory of mental models, it can be thought that, if (1), (2), (3), and (4) were taken as elements to the construction of scientific language, that would not require additional greater intellectual effort from

[3] Carnap, "Testability and meaning," 442.
[4] Carnap, "Testability and meaning," 442.
[5] Carnap, "Testability and meaning," 442.
[6] Carnap, "Testability and meaning," 442.
[7] Carnap, "Testability and meaning," 442-443.
[8] E.g., Ruth M. J. Byrne and Philip N. Johnson-Laird, "*If* and *or*: Real and counterfactual possibilities in their truth and probability," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46, 4 (2020): 760-780.

researchers will be pointed out. Thus, it will also be argued that, beyond Carnap's initial intentions, on the contrary, the limitations of those formulae would lead to simplify their cognitive activity.

Four Formulae and Three Rules for the Building of Scientific Language

Carnap's proposal[9] is much more comprehensive and is not limited to the four formulae above. In fact, (1), (2), (3), and (4) are examined by Carnap with much more attention than the way they will be dealt with here. Nevertheless, that is not a problem because the aim of this paper is simple. It is just to show that the formulae and their restrictions, at a minimum, it their initial presentation, do not add further requirements to researchers' activity. In fact, they can even make that very activity easier.

Focusing on the formulae, if the following equivalences are assumed, it is not difficult to get an example for (1) with thematic content.

Q_1: To be an animal.

Q_2: To be rational.

Q_3: To be a human being.

Thus, the example would be:

(5) For every x, IF x is an animal THEN (IF x is rational THEN x is a human being)

Nonetheless, according to Carnap,[10] (1) should be a 'reduction sentence' regarding its last consequent, that is, Q_3 (or 'to be a human being'). This means that (1) should be a sentence related to the degree in which a predicate such as Q_3 can be confirmed.

At this point, perhaps it is important to remind that, although that is sometimes forgotten, Carnap's view in this connection is not very different from the one of Popper.[11] This is explicitly mentioned by Carnap.[12] In this sense, it seems that Carnap's idea of reduction does not necessarily refer to definitive confirmation. In a similar manner as for Popper, it can be related to just levels of confirmation.

But, this said, the most relevant point here is that, for Carnap,[13] (1) can be a reduction sentence for the consequent of the conditional between its brackets if and only if a condition is fulfilled:

---

[9] Carnap, "Testability and meaning," 419-471.

[10] Carnap, "Testability and meaning," 442.

[11] Karl Popper, *The Logic of Scientific Discovery* (New York: Routledge Classics, 2002).

[12] Carnap, "Testability and meaning," 426.

[13] Carnap, "Testability and meaning," 442.

(6) Not-(Q_1 & Q_2) cannot be a valid formula.

Undoubtedly, '&' in (6) indicates logical conjunction, and, because (1) is universally quantified, the meaning of (6) is obvious. (1) can be admitted if and only if both Q_1 and Q_2 can be true, the cases in which one of those predicates is false having to be ignored.

This is not a difficulty for example (5) at all. (5) can be a reduction sentence for the predicate 'to be a human being' (Q_3) because there are animals (Q_1) and there are rational beings (Q_2). The point of (6) for the goals of this paper will be made below.

As far as (2) and (3) are concerned, the previous definitions of Q_1, Q_2, and Q_3 can be kept to offer examples with thematic content. It would only be necessary to add these new equivalences:

Q_4: To be a mammal.

Q_5: To have fins.

Thereby, the example for (2) would be the same as the one for (1), that is, (5). The example for (3) would be as follows:

(7) For every x, IF x is a mammal THEN (IF x has fins THEN x is not a human being)

However, there is a condition given by Carnap[14] here too. (2) and (3) can be a 'reduction pair' for the last predicate in (2), or the predicate negated in (3), that is, again, Q_3, if and only if:

(8) Not-[(Q_1 & Q_2) OR (Q_4 & Q_5)] cannot be valid.

'OR' denotes in (8) logical inclusive disjunction. Therefore, what (8) implies is, in a similar way as in the previous case, that (2) and (3) can be assumed if and only if Q_1, Q_2, Q_3, and Q_4 can be all true. The situations in which that does not happen are irrelevant and should not be considered. Nevertheless, examples (5) and (7) continue not to be a problem. As said, there are animals and rational beings, but there are both mammals (Q_4) and animals having fins (Q_5) too. The importance of (8) will be made explicit below as well.

Finally, with regard to (4), Carnap[15] says that it is a special case in which these equivalences occur:

$Q_4 = Q_1$

$Q_5 = \text{not-}Q_2$

[14] Carnap, "Testability and meaning," 442.

[15] Carnap, "Testability and meaning," 441-442.

Given these last equivalences, following Carnap,[16] it can be stated that (3) could be transformed into:

(9) IF Q_1 THEN (IF not-Q_2 THEN not-Q_3)

As Carnap[17] reminds, (9) is equivalent to:

(10) IF Q_1 THEN (IF Q_3 THEN Q_2)

And, as he also points out,[18] from (1) and (10), it is possible to derive (4).

It is easy to give an example with thematic content for (4) from the previous equivalences too:

(11) For every x, IF x is an animal THEN (x is a human being IFF x is rational)

Nonetheless, (4) has, following Carnap,[19] its restriction as well. (4) is a 'bilateral reduction sentence' for the left predicate of the biconditional in its brackets, that is, again, for Q_3, if and only if this is correct:

(12) For every x, not-Q_1 cannot be valid.

Because of the previous definition of Q_1, that is, 'to be an animal,' (12) is not a difficulty for (11) either. What (12) provides is the need for Q_1 to be possible, since it leads not to take into account the cases in which Q_1 does not happen. Nevertheless, none of this has an influence on (11), since, as said, there are animals.

A theory such as the theory of mental models can show the interest that restrictions such as (6), (8), and (12) can have, irrespective of Carnap's real perspective, from the cognitive point of view. Those restrictions can in turn cause (1), (2), (3), and (4) to be very relevant elements in the process of construction of scientific language. But to show why all of this is the case, it is necessary to explain some theses of the theory of mental models before.

The Conjunction of Possibilities of the Conditional

For the theory of mental models, the conditional is a 'sentential connective.'[20] Of course, the conditional is not the only sentential connective the theory of mental

[16] Carnap, "Testability and meaning," 442.
[17] Carnap, "Testability and meaning," 442.
[18] Carnap, "Testability and meaning," 442.
[19] Carnap, "Testability and meaning," 443.
[20] E.g., Philip N. Johnson-Laird and Marco Ragni, "Possibilities as the foundation of reasoning," *Cognition* 193 (2019). DOI: 10.1016/j.cognition.2019.04.019.

models addresses.[21] However, for the aims of the present paper, the conditional is the most important connective to deal with.

Sentential connectives, including the conditional, link sentences to 'conjunctions of possibilities.'[22] Thus, in the case of a conditional such as:

(13) IF Q_2 THEN Q_3

Following the usual way to express conjunctions of possibilities in the latest papers supporting the theory,[23] its conjunction of possibilities would be akin to this one:

(14) Possible (Q_2 & Q_3) & Possible (not-Q_2 & Q_3) & Possible (not-Q_2 & not-Q_3)

Nevertheless, another important aspect of the theory of mental models is that its proponents often deem it as a 'dual-process' theory.[24] As it is well known, dual-process theories[25] distinguish between two systems. Those systems are usually named 'System 1' and 'System 2.' Generally, System 1 allows inferring quick conclusions because it does not carry out deductive processes. Basically, it is the system responsible for the activities related to intuition. However, System 2 works in a slow way. This is because it leads deductive processes. Accordingly, it can be said that its mental processes are more rigorous.

Under this framework, the theory of mental models also claims that individuals might use only one of those two systems. This, in the case of the conditional, means that, when the system working is System 1, only the first conjunct in a conjunction of possibilities such as (14) is detected. The other two

[21] E.g., Philip N. Johnson-Laird, "Inference with mental models," in *The Oxford Handbook of Thinking and Reasoning*, eds. Keith J. Holyoak and Robert G. Morrison (New York: Oxford University Press, 2012), 134-145.

[22] See also, e.g., Sangeet Khemlani, Thomas Hinterecker, and Philip N. Johnson-Laird, "The provenance of modal inference," in *Computational Foundations of Cognition*, eds. Glenn Gunzelmann, Andrew Howes, Thora Tenbrink, and Eddy J. Davelaar (Austin: Cognitive Science Society, 2017), 663-668.

[23] E.g., Johnson-Laird and Ragni, "Possibilities as the foundation of reasoning"; Khemlani et al., "The provenance of modal inference," 663-668.

[24] E.g., Philip N. Johnson-Laird, Sangeet Khemlani, and Geoffrey P. Goodwin, "Logic, probability, and human reasoning," *Trends in Cognitive Sciences* 19, 4 (2015): 201-214.

[25] E.g., Jonathan St. B. T. Evans, "How many dual-process theories do we need? One, two or many?" in *In Two Minds: Dual Processes and Beyond*, eds. Jonathan St. B. T. Evans and Keith Frankish (Oxford: Oxford University Press, 2009), 33-54; Keith Stanovich, "On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning," in *The Oxford Handbook of Thinking and Reasoning*, eds. Keith J. Holyoak and Robert G. Morrison (New York: Oxford University Press, 2012), 343-365.

conjuncts are harder to identify and require the action of System 2.[26] In fact, the theory of mental models deems the second and third conjuncts as presuppositions people make when they manage to note that those conjuncts are possibilities for a conditional.[27]

From this point of view, it can be thought that any activity demanding only to identify the first conjunct and hence to use System 1 must be a not challenging activity. This is what occurs with Carnap's rules and it is shown in the next section.

Reduction and System 1

Indeed, if the theory of mental models is right, it seems that what is proposed by Carnap[28] is not really a great requirement for scientists or philosophers of science. In this way, there are many reasons why the theory should be accepted. Its predictions have been confirmed many times.[29] Besides, its proponents have developed even a software (mReasoner) which, following strictly the main principles of the theory, tries to imitate human reasoning.[30] However, beyond that discussion, what will be argued below, as said, is just that, if the theory of mental models is correct, the mental activity demanded by Carnap[31] to build scientific language is very simple and basic, since it only needs to use System 1.

Starting by (1), it can be claimed that, if only System 1 were utilized, its only possibility would be:

(15) Possible [Q_1 & (IF Q_2 THEN Q_3)]

Nevertheless, if people resorted to System 2 instead, two more conjuncts should be added:

(16) Possible [Q_1 & (IF Q_2 THEN Q_3)] & Possible [not-Q_1 & (IF Q_2 THEN Q_3)] &

[26] See also, e.g., Johnson-Laird, "Inference with mental models," 134-145.

[27] E.g., Johnson-Laird and Ragni, "Possibilities as the foundation of reasoning."

[28] Carnap, "Testability and meaning," 419-471.

[29] See, in addition to the works supporting the theory of mental models cited along the present paper, e.g., Monica Bucciarelli and Philip N. Johnson-Laird, "Deontics: Meaning, reasoning, and emotion," *Materiali per una Storia della Cultura Guiridica* XLIX, 1 (2019): 89-112; Sangeet Khemlani and Philip N. Johnson-Laird, "Why machines don't (yet) reason like people," *Künstliche Intelligenz* 33 (2019): 219-228; Ana Cristina Quelhas, Célia Rasga, and Philip N. Johnson-Laird, "The analytic truth and falsity of disjunctions," *Cognitive Science* 43, 9 (2019). DOI: https://doi.org/10.1111/cogs.12739.

[30] See, e.g., Khemlani et al., "The provenance of modal inference," 663-668; Johnson-Laird et al., "Logic, probability, and human reasoning," 201-214; for download: https://www.modeltheory.org/models/mreasoner/

[31] Carnap, "Testability and meaning," 419-471.

Possible [not-Q_1 & not-(IF Q_2 THEN Q_3)]

But the second conjunct in all of the possibilities both in (15) and (16) is also a conditional (IF Q_2 THEN Q_3). So, that conjunct would be linked to possibilities too, and (15) and (16) would not be the final conjunctions of possibilities. In the case of (15), that is, of the situation in which only System 1 is used, the second conjunct of the possibility has to be transformed in this way:

(17) Possible [Q_1 & Possible (Q_2 & Q_3)]

Certainly, the conditional (IF Q_2 THEN Q_3) is again transformed into just a possibility [Possible (Q_2 & Q_3)], since, as indicated, when the system that is taken into account is System 1, only the first conjunct in a conjunction of possibilities such as (14) is considered. However, if so, (17) is describing only a possible scenario, which can be better expressed as follows:

(18) Possible (Q_1 & Q_2 & Q_3)

Undoubtedly, the situation is much more complex when the system is System 2. In that circumstance, the conditional (IF Q_2 THEN Q_3) has to be deployed in three possibilities in the cases of the two first conjuncts in (16). That is not necessary for the last conjunct, as it is negated. Nonetheless, that negation implies an additional difficulty too. It requires to understand that the negation of a conditional refers to the affirmation of its antecedent and the negation of its consequent. Thus, the conjunction of possibilities in the case of System 2 would be:

(19) Possible [Q_1 & Possible (Q_2 & Q_3)] & Possible [Q_1 & Possible (not-Q_2 & Q_3)] & Possible [Q_1 & Possible (not-Q_2 & not-Q_3)] & Possible [not-Q_1 & Possible (Q_2 & Q_3)] & Possible [not-Q_1 & Possible (not-Q_2 & Q_3)] & Possible [not-Q_1 & Possible (not-Q_2 & not-Q_3)] & Possible [not-Q_1 & Possible (Q_2 & not-Q_3)]

Of course, (19) can be simplified if its possibilities are manipulated in a way similar to the one used to transform (17) into (18). Thereby, the result would be:

(20) Possible (Q_1 & Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & Q_2 & not-Q_3)

Still, (20) keeps being a very complex conjunction of possibilities. And it could be even harder to manage. The theory of mental models also raises the fact that people can tend to negate a conditional in a manner different from that suitable in classical logic. They can interpret that the negation of a sentence such as (13) is equivalent to that very sentence with its consequent negated,[32] that is, to:

[32] See, e.g., Sangeet Khemlani, Isabel Orenes, and Philip N. Johnson-Laird, "The negation of

(21) IF Q_2 THEN not-Q_3

And System (2) would lead to these possibilities for (21):

(22) Possible (Q_2 & not-Q_3) & Possible (not-Q_2 & not-Q_3) & Possible (not-Q_2 & Q_3)

Which in turn would increase the number of conjuncts in (20). It would be necessary to consider nine possibilities, as shown in (23).

(23) Possible (Q_1 & Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & not-Q_3) & Possible [not-Q_1 & Possible (Q_2 & not-Q_3)] & Possible [not-Q_1 & Possible (not-Q_2 & not-Q_3)] & Possible [not-Q_1 & Possible (not-Q_2 & Q_3)]

Or, if preferred, by simplifying the three last conjuncts in (23) as done in (18) from (17) and in (20) from (19):

(24) Possible (Q_1 & Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & Q_3) & Possible (Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & Q_3) & Possible (not-Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & Q_2 & not-Q_3) & Possible (not-Q_1 & not-Q_2 & not-Q_3) & Possible (not-Q_1 & not-Q_2 & Q_3)

However, at this point, what is important is that the rule had a limitation: (6). If (6) were followed, all of the possibilities in (24) including not-Q_1 or not-Q_2 would have to be removed. But, if that were done, the result would be (18), and only System 1 is needed to come to (18). Accordingly, Carnap's[33] proposal regarding (1) does not imply a greater mental effort. Under the theses of the theory of mental models, the cognitive effort to work with (1) would be minimal, since eight of the nine possibilities could be ignored. Furthermore, for the theory, the eight possibilities that would not be necessary to take into account are the possibilities related to higher levels of cognitive difficulty.

Something similar can be argued with regard to (2) and (3). Because it is identical to (1), the case of (2) can be considered already explained. As far as (3) is concerned, its account would be easy to present too.

Based upon the previous explanation, if only System 1 were used, (3) would have just a possibility:

(25) Possible (Q_4 & Q_5 & not-Q_3)

Nevertheless, paying attention to System 2, six more possibilities would have to be added:

conjunctions, conditionals, and disjunctions," *Acta Psychologica* 151 (2014): 1-7.

[33] Carnap, "Testability and meaning."

(26) Possible (Q_4 & Q_5 & not-Q_3) & Possible (Q_4 & not-Q_5 & not-Q_3) & Possible (Q_4 & not-Q_5 & Q_3) & Possible (not-Q_4 & Q_5 & not-Q_3) & Possible (not-Q_4 & not-Q_5 & not-Q_3) & Possible (not-Q_4 & not-Q_5 & Q_3) & Possible (not-Q_4 & Q_5 & Q_3)

As in the previous case, the number of possibilities can be even greater. If the trend to understand negated conditionals to be conditionals in which just its consequent is negated is assumed here as well, the real conjunction of possibilities may not be (26), but (27).

(27) Possible (Q_4 & Q_5 & not-Q_3) & Possible (Q_4 & not-Q_5 & not-Q_3) & Possible (Q_4 & not-Q_5 & Q_3) & Possible (not-Q_4 & Q_5 & not-Q_3) & Possible (not-Q_4 & not-Q_5 & not-Q_3) & Possible (not-Q_4 & not-Q_5 & Q_3) & Possible (not-Q_4 & Q_5 & Q_3) & Possible (not-Q_4 & not-Q_5 & Q_3) & Possible (not-Q_4 & not-Q_5 & not-Q_3)

Nonetheless, in any case, the restriction (8) also simplifies the number of conjunctions here. (8) includes a negated disjunction, and the first disjunct of that disjunction (Q_1 & Q_2) can play exactly the same role as (6) for (24) in the case of (2). As pointed out, (2) is identical to (1), and, therefore, the explanation for (1) also holds for (2). So, as (1), (2) would only admit (18) and, as also indicated, (18) can be detected resorting only to System 1. On the other hand, the second disjunct (Q_4 & Q_5) of that very disjunction in (8) prevents from considering possibilities with not-Q_4 or not-Q_5, which leads to (25). But, as accounted for too, (25) is the possibility that could be derived from (3) by means of just System 1. Again, this demonstrates that no special mental effort is required for this rule either.

Hence, only the case of (4) remains to explain. It is true that the theory of mental models also offers an account of the biconditional that is not exactly identical to the one it gives for the conditional.[34] Nevertheless, given that the expression 'if and only if' is not very usual in natural language, that, in many cases, biconditionals are really expressed with the form of conditional sentences, that is, using the words 'if' and 'then,'[35] and that people generally come to biconditional interpretations from conditional sentences by considering the second possibility in conjunctions such as (14) to be unacceptable,[36] perhaps it is better that the explanation of (4) is not based on that account. It can be easier just to focus on the fact that, according to Carnap,[37] as indicated, (4) is actually the result of combining (1) and (10).

Thereby, to analyze (4) it can be enough to review, separately, (1) and (10) in order to check if System 2 is necessary in their cases. (1) has already been dealt with.

[34] E.g., Johnson-Laird, "Inference with mental models," 134-145.

[35] E.g., Philip N. Johnson-Laird and Ruth M. J. Byrne, "Conditionals: A theory of meaning, pragmatics, and inference," *Psychological Review* 109, 4 (2002): 646-678.

[36] See also, e.g., Johnson-Laird and Byrne, "Conditionals," 646-678.

[37] Carnap, "Testability and meaning," 442.

As seen, given restriction (6), only (18) needs to be addressed in its particular case. But, as argued, System 1 suffices to come to (18).

As far as (10) is concerned, it is evident that its explanation should not be very different from that of (1). If only System 1 works, only this possibility can be assigned to (10):

(28) Possible (Q_1 & Q_3 & Q_2)

As it can be noted, (28) is identical to (18). Just the order of the conjuncts into the possibility is different between them.

On the other hand, when (10) is processed by System 2, the conjunction of possibilities is this one:

(29) Possible (Q_1 & Q_3 & Q_2) & Possible (Q_1 & not-Q_3 & Q_2) & Possible (Q_1 & not-Q_3 & not-Q_2) & Possible (not-Q_1 & Q_3 & Q_2) & Possible (not-Q_1 & not-Q_3 & Q_2) & Possible (not-Q_1 & not-Q_3 & not-Q_2) & Possible (not-Q_1 & Q_3 & not-Q_2)

Of course, here it can also be interpreted that conditionals are negated by simply negating their consequents. That would allow transforming (29) into (30).

(30) Possible (Q_1 & Q_3 & Q_2) & Possible (Q_1 & not-Q_3 & Q_2) & Possible (Q_1 & not-Q_3 & not-Q_2) & Possible (not-Q_1 & Q_3 & Q_2) & Possible (not-Q_1 & not-Q_3 & Q_2) & Possible (not-Q_1 & not-Q_3 & not-Q_2) & Possible (not-Q_1 & Q_3 & not-Q_2) & Possible (not-Q_1 & not-Q_3 & not-Q_2) & Possible (not-Q_1 & not-Q_3 & Q_2)

But, once again, what is important is the restriction. Because, as argued by Carnap,[38] (4) comes from a special case of (3) in which $Q_4 = Q_1$ and Q_5 = not-Q_2, one might think that, beyond (12), the restriction for (3), that is, (8) applies to (10) too. This is even less difficult to accept if it is noted that (8) is stronger than (12). Thus, the first disjunct in the disjunction in (8) allows eliminating the cases of not-Q_1 and not-Q_2 in both (29) and (30), which leads to (31).

(31) Possible (Q_1 & Q_3 & Q_2) & Possible (Q_1 & not-Q_3 & Q_2)

So, the problem would be only the second possibility in (31). It is a possibility for (10) that requires System 2 to be detected and, therefore, seems to undermine the thesis of the present paper. Nevertheless, this is not necessarily that way. Even if System 2 were used, the second conjunct in (31) would be never considered. If an individual resorts to System 2, this last system can note that, for a sentence such as (13), the only forbidden possibility is that missing in (14), that is, that in which Q_2 happens and Q_3 does not occur. A situation such as this one is exactly what is described in the second possibility of (31). In this manner, because (4) is built by

[38] Carnap, "Testability and meaning," 441-442.

means of (1) and (10), and (13) is included in (1), the second conjunct in (31) could be attributed to (4) in no way, whether or not System 2 is used.

That means that only the first possibility in (31) is relevant for (4), and, as indicated, that is the possibility corresponding to (10) if only System 1 is taken into account - it is equivalent to (28). Accordingly, it can be said that the possibilities that could be detected for (4) by resorting to System (2) would not be suitable for it either.

Therefore, none of the rules addressed in this paper would require the effort related to System 2. In all the cases, because of restrictions, it would be enough to use System 1.

Conclusions

This reveals that, as said, the particular instructions given by Carnap[39] to build scientific language dealt with here do not necessarily make scientific or philosophical work harder (regardless of Carnap's real intentions with those instructions). This is true, at least, from the framework of the theory of mental models.

It is not the first time the theory of mental models is linked to Carnap's approach. The literature shows several examples. However, most of the links are provided to his semantic method of extension and intension.[40] So, such links can appear to be obvious, as an essential element of that method is a set of 'state-descriptions' or 'possible worlds.'[41]

But, beyond those facts, perhaps it is important to highlight that several points remain to be explored. The paper by Carnap[42] not only analyzes formulae (1), (2), (3), and (4). It also reviews much more different aspects related to language, science, testability, and confirmation. The paper has even a second part that keeps moving forward from his ideas.[43] Accordingly, maybe it would be relevant to continue to study the greatest possible number of theses of Carnap's general work in order to check whether or not their trend is not to imply additional effort to scientific tasks and activities regarding the development and application of knowledge.

On the other hand, Carnap's theoretical framework should be somehow reviewed by means of its implementation in practice as well. Thus, it would be

[39] Carnap, "Testability and meaning," 419-471.

[40] Rudolf Carnap, *Meaning and Necessity: A Study in Semantics and Modal Logic* (Chicago: The University of Chicago Press, 1947).

[41] For some relations between these two frameworks, see, e.g., Miguel López-Astorga, "Apparent L-falsity and actual logical structures," *Problemos* 97 (2020): 114-122.

[42] Carnap, "Testability and meaning," 419-471.

[43] Rudolf Carnap, "Testability and meaning – Continued," *Philosophy of Science* 4, 1 (1937): 1-40.

opportune to confirm that formulae such as (1), (2), (3), and (4), and restrictions such as (6), (8), and (12) can be used in very different scientific fields and capture their concepts and theses. It is evident that this would be a gradual work of confirmations that, probably, it would never finish. However, it would not be actually very far from Carnap's approach. As mentioned, in this particular aspect, his view is not very different from Popperian philosophy. For that, it would be coherent with his ideas to keep researching in this direction.

In any case, what can be stated for sure currently is that, if the theory of mental models can be accepted, particular proposals by Carnap such as those analyzed above can continue to be interesting. As also stated, there are reasons for assuming the theory of mental models and, under its approach, the rules reviewed can be advisable. In addition to give clarity to the daily work carried out by scientists, and offer relevant inputs for the debate about the characteristics that scientific language should have, they do not lead to higher levels of effort from the cognitive point of view. All of this, of course, beyond Carnap's actual goals.[44]

[44] **Acknowledgments:** PIA Ciencias Cognitivas, Centro de Investigación en Ciencias Cognitivas, Instituto de Estudios Humanísticos, Universidad de Talca. Fondo Fondecyt de Continuidad para Investigadores Senior, código FCSEN2102, Universidad de Talca.

ON THE VERY IDEA OF UNDERCUTTING DEFEAT

Erhan DEMIRCIOGLU

ABSTRACT: My aim in this paper is to cast doubt on the idea of undercutting defeat by showing that it is beset by some serious problems. I examine a number of attempts to specify the conditions for undercutting defeat and find them to be defective. Absent further attempts, and on the basis of the considerations offered, I conclude that an adequate notion of undercutting defeat is lacking.

KEYWORDS: evidential defeat, undercutting defeat, rebutting defeat, evidential support, misleading evidence

1. Undercutting Defeat: "The Unpolished Conception"

In order for a given case to be a case of evidential defeat, the following conditions (or their respective instantiations)[1] must simultaneously obtain:

(i) The subject *S* has a piece of evidence *E* that supports *p*.

(ii) The subject *S* has a piece of evidence *E´* such that *E* and *E´* together do not support *p*.[2]

A widely recognized distinction, famously drawn by Pollock,[3] is between two kinds of evidential defeat, viz. between rebutting and undercutting defeat. The distinction depends on two different ways of satisfying (ii). Rebutting defeat occurs when a subject has some evidence that supports a proposition and also has some other evidence that supports the negation of that proposition, in which case the subject's

[1] For convenience's sake, I sometimes drop this otherwise important qualification below and pretend that such variables as *E* and *p* are constants. This gives rise to an innocuous shifting between talk of truth and talk of satisfiability (e.g., between the truth of (i) and the satisfiability of (i)).

[2] Conditions (i) and (ii) together define what one might call "total defeat," where E´ *totally* defeats the support *E* provides to *p*. There is also what one might call "partial defeat," where *E partly* defeats the support *E* provides to *p*. In what follows, nothing substantive hangs on the distinction between total and partial defeat and the discussion will proceed on the definition of (total) evidential defeat provided.

[3] J. L. Pollock, *Contemporary Theories of Knowledge* (Totowa, NJ: Roman and Littlefield, 1986).

total evidence does not support the proposition in question. In a case of rebutting defeat, the following condition obtains:

(iii*) The subject has a piece of evidence E' that supports not-p.

It is the joint truth of (i) and (iii*), which one might call "counterbalancing," that makes a given case a case of *rebutting* defeat. Furthermore, counterbalancing guarantees the truth of (ii), which combined with (i), entails that a given case is a case of *evidential* defeat.

Now, here is a paradigm case presented as involving undercutting defeat in the literature. Entering my friend's reading room, I see some books that appear red. On this basis, I form the belief that there are red books in the room. My friend then informs me that the books are intricately illuminated by red light. Despite the fact that my friend's testimony is *not* evidence that there are no red books in the room, my total evidence does not support the belief that there are red books in the room. In this purported case of undercutting defeat, my friend's testimony "attacks"[4] and "severs"[5] the evidential connection between my experience as of seeing red books and the belief that there are red books (rather than "attacking [the belief] itself"[6]). So, given my friend's testimony, my experience is not "an indication of the truth of [the belief]."[7] The support my experience provides to the belief is destroyed, as it were, by my friend's testimony. In a case of undercutting defeat, then, the following condition is suggested by the foregoing remarks:

(iii**) The subject has a piece of evidence E' that destroys the support E provides to p.

It is, it is claimed, the joint truth of (i) and (iii*), which one might call "destruction," that makes a given case a case of *undercutting* defeat. Furthermore, destruction guarantees the truth of (ii), which combined with (i), entails that a given case is a case of *evidential* defeat.

The notion of undercutting defeat is underdeveloped because there is no account available in the literature of what it is to destroy a given evidential support (or, equivalently, what it is to sever an evidential connection). It is clear that what is meant by "destruction" (or "severing") cannot simply be the joint truth of (i) and (ii) since that does not distinguish undercutting from rebutting defeat. Then, what

[4] Pollock, *Contemporary Theories of Knowledge*, 196.

[5] Thomas Kelly. "Evidence," in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2016, URL: https://plato.stanford.edu/archives/win2016/entries/evidence/

[6] Pollock, *Contemporary Theories of Knowledge*, 196.

[7] Stewart Cohen, "Justification and Truth," *Philosophical Studies* 46 (1984): 279-295, 290.

exactly does "destruction" mean here? A satisfactory answer to this question is curiously lacking.

However, there is a glaring and more serious, albeit as we shall see related, problem I would like to point out about the characterization above of undercutting defeat. The problem is that (i) and (iii**) cannot be simultaneously true: if *E´* destroys the support *E* provides to *p*, then *E* does not support *p*, and *vice versa*. So, if what is required for a case to be a case of undercutting defeat is the joint truth of (i) and (iii**), then there is no such thing as undercutting defeat. Undercutting defeat appears to demand an impossible trick to be accomplished: you cannot have your cake and eat it too, and you cannot have *E* supporting *p* and *E´* destroying the support *E* provides to *p* too. And, note that rebutting defeat, which requires counterbalancing but not destruction, faces no analogous threat.

Let me call the conception of undercutting defeat according to which undercutting defeat requires the joint truth of (i) and (iii**) "the unpolished conception." I take it that the argument just offered conclusively shows that the unpolished conception of undercutting defeat is inconsistent. Of course, the intention here is not to attribute something as overtly problematic as the unpolished conception to such noteworthy advocates of undercutting defeat as Pollock and others. Still, my impression is that the unpolished conception is not entirely off the mark and it is at least suggested by some unguarded definitions of undercutting defeat. And, more significantly, the failure of the unpolished conception brings explicitly into view the task of providing an adequate conception of undercutting defeat, one in which something along the lines of destruction of evidential support plays a central role. Destruction or something like it is what distinguishes undercutting from rebutting defeat. However, the problem is that evidential support destruction seems to guarantee that a condition for evidential defeat (namely, (i)) is not satisfied. So, the challenge is this: how can there be a case of evidential defeat in which evidential support destruction plays a central role?

Can the challenge be met? Clearly, either (i) or (iii**) (or both) must go and replaced by some other condition, in a way that respects the idea of evidential support destruction. In the next section, I will consider a readily available attempt along these lines and argue that it does not work either.

2. "The Diachronic Conception" Considered

According to what I shall call the "diachronic" conception, the following conditions must obtain for a given case in order for it to be a case of undercutting defeat:

(i1) At t_1, *S* has a piece of evidence *E* that supports *p*.

(iii1) At a later time t_2, *S* acquires a piece of evidence *E´*, which destroys the support

that *E* (previously) provided to *p*.

The diachronic conception holds that the duo of (i_1) and (iii_1) is consistent.[8]

I want to make two points about the diachronic conception, one short and the other longer. As for the former, let us start with observing that the qualification "later" in (iii_1) is essential: the diachronic conception is consistent only if t_2 is not the same time as t_1. More specifically, if t_1 and t_2 were the same time, then the diachronic conception would suffer from the same inconsistency problem that we have seen afflicts the unpolished conception. Now, and this is my short point, the diachronic conception cannot account for some cases that, if there is such a thing as undercutting defeat, should qualify as cases of undercutting defeat. Consider, for instance, a slightly revised version of "the red lighting case" presented in the previous section. In this version, let us assume, I acquire both pieces of evidence *at the same time*: the time when I see some books that appear red is the same as the time when my friend informs me that they are illuminated by red light. If there is such a thing as undercutting defeat, then this case is presumably an example of undercutting defeat. However, the diachronic conception entails that it is not (because (iii_1) does not obtain). So, the diachronic conception is inadequate.

The longer point I wish to make will take some time to develop. First, here is a note about how (iii_1) is to be understood. As it stands, (iii_1) is ambiguous between two different readings: is what destroys the (previous) support *E* provides to *p*, the evidence *E´* itself or the subject's acquisition of *E* ? Let us reconsider the red lighting case. Is what is supposed to destroy the support my experience provides to the belief that there are red books around, my friend's testimony itself or my 'acquisition' (or 'awareness' (or whatever is required for possessing evidence)) of that testimony? It is clear that under those circumstances in which my friend's testimony is present while I am unaware of it (if, for instance, my friend is only engaged in soliloquy and says that the books are illuminated by red light 'all too silently'), the support my experience provides to the belief in question must remain intact. So, the support my experience provides to the belief is not destroyed when my friend says what he does but when I become aware of what he says. It is my *acquisition* of my friend's testimony, and not merely the testimony itself, that is supposed to have the destructive effect on the support my experience provides to the belief.[9] The

[8] The diachronic conception might also replace (ii) by (ii_1), which reads: at a later time t_2, *S* acquires a piece of evidence *E´* such that *E* and *E´* together do not support *p*.

[9] Of course, there must be another side to the story: if my friend's testimony did not have the *potential* to destroy the support in question, then my acquisition of it would not destroy it. One way to put the point is, then, this: what explains the fact that my acquisition of the testimony has the putative destructive effect is that the destructive potential of my friend's testimony is

diachronic conception assumes, then, what I will call "the destructibility thesis," according to which a subject's acquisition of a piece of evidence might destroy the support another piece of evidence provides to a proposition. If the destructibility thesis is false, then the diachronic conception fails.

In the remainder of this section, I will argue that the destructibility thesis is false. Before that, however, let me note why the destructibility thesis might appear to be true. I grant that the main idea behind the diachronic interpretation is intuitive: a subject can have E that supports p *until* a certain time (viz., the time when she acquires E' which destroys the support). Before the acquisition of E', E supports p. After its acquisition, however, it *no longer* does. Here is an analogy. Think about "the support" at a certain time a wife provides to her husband. At a later time, the wife learns that he is cheating on her. After that time, she no longer supports her. We can say that the wife's learning about her husband's disloyalty "destroys" the support she previously provided to her. Similar considerations might appear to apply to evidential support relations: the support E provides to p might get destroyed when the subject acquires E' – or so it is held by the diachronic interpretation. Why not say, for instance, that my experience as of seeing red books in the room supports the belief that there are red books there *until* my friend informs me about the red lighting, and after that, it does not because my friend's testimony *destroys* the previous support?

Whatever its intuitive appeal might be, however, the destructibility thesis is false. First, note that there is a good reason to be suspicious of the analogy between evidential support relations and "wifey" support relations. When the wife learns that her husband is cheating on her, she goes through a certain change (e.g., she now believes that his husband is cheating on her, she is disappointed and angry with him, and so on). This change explains why she *no longer* supports her. However, when a subject acquires E', E does not go through any change. True, the subject herself goes through a certain change (more specifically, her total evidence changes); but, the relevant point is that E is the same as before (and this is the relevant point because the relevant question is whether E continues to support p after the subject acquires E' (and not whether her total evidence does)). And, if E supports p before the subject acquires E', and also if E does not go through any change after the subject acquires E', then how can E cease to support p after the subject acquires E'? How can the subject's acquisition of E' possibly destroy the support E provides to p, if E remains the same as before?

---

actualized by my acquisition of it. Despite this, however, the main point stands that it is, on the diachronic conception, my acquisition of the testimony that has the putative destructive effect.

Having noted a good reason to think that the analogy breaks down between evidential and wifey support relations, here is now a direct consideration against the destructibility thesis. It is clear that if *E* supports *p*, then *E* supports *p* whether I, or any other subject, has *E*. And, it is equally clear that if *E* does not support *p*, then *E* does not support *p* whether I, or any other subject, has *E*. That is, evidential support relations hold independently of how pieces of evidence are distributed among subjects, which I will call "distributional independence." It seems clear that what a given piece of evidence supports does not depend on *who* has that evidence or even on whether *anyone* has that evidence or not. The question "does *E* support *p*?" can be answered without raising the question "who has *E*?" or "does anyone have *E*?" And, note that rejecting distributional independence entails that by *re*distributing pieces of evidence among the subjects, we can change evidential support relations that hold between those pieces of evidence. And, since, as I believe many would agree, that cannot be right, distributional independence must be accepted.

However, distributional independence is violated by the destructibility thesis: if a subject's *acquisition* of a piece of evidence can have a destructive effect on an evidential support relation, as the destructibility thesis says it can, then contra distributional independence, that relation is not independent of how evidence is distributed among subjects. Given distributional independence, then, the destructibility thesis is false and evidential support relations cannot be destroyed by the *acquisition* of further piece of evidence.

I maintain that the argument from distributional independence just rehearsed shows that the destructibility thesis must be rejected. However, it might be argued that the diachronic conception does not need to assume a thesis as strong as the destructibility thesis, that a weaker version of the destructibility thesis would do. According to what we might call "the weak-destructibility thesis," a subject's acquisition of a piece of evidence might destroy *for her* (whatever this might plausibly mean, and as opposed to *in general*) the support another piece of evidence provides to a proposition. Since the weak-destructibility thesis is consistent with distributional independence, the response goes, the diachronic conception resting on the former is not threatened by an argument from the latter.

I agree that the weak-indestructibility thesis is consistent with distributional independence: in particular, the idea that a given evidential support relation might be destroyed *for a subject* by her acquiring further pieces of evidence is consistent with the fact that evidential support relations hold *in general* independently of how pieces of evidence are distributed among subjects. However, the weak-indestructibility thesis and distributional independence together leads to an absurdity. Suppose that everyone acquires E', in which case the evidential support

relation between *E* and *p* is *destroyed for everyone*. However, distributional independence entails that the evidential support relation between *E* and *p* goes undestroyed by the fact that everyone acquires *E´* and hence that it continues to *hold in general*. However, this is absurd: what does it mean to say that the evidential support relation between *E* and *P* is destroyed for everyone, despite the fact that it continues to hold in general? Since I believe distributional independence is true, the weak-destructibility thesis must be rejected in order to avoid the absurdity.

On the basis of the considerations above, I maintain that both the destructibility thesis and the weak-indestructibility thesis are false. And, if they are false, the diachronic conception that rests on one or the other fails.

I would like to conclude this section with two clarificatory remarks. First, my point against the destructibility thesis does not cast any doubt on rebutting defeat (or evidential defeat in general). This is because rebutting defeat does not require that evidential support relations be destroyable by the acquisition of further pieces of evidence. Second, I don't deny that there might be a sense in which evidential support relations might be destroyed. If, for instance, that *E* supports *p* is a contingent fact, then *E* might support *p* at a certain time and not support it at a different time. My point is, to emphasize, merely that evidential support relations cannot be destroyed by the acquisition of further pieces of evidence.

3. Other Conceptions Considered

We have seen that neither of the two conceptions – namely, the unpolished conception and the diachronic conception – that can be more or less directly gleaned from the literature on undercutting defeat works. I will now proceed to assessing some other possible conceptions of undercutting defeat, which might further be offered as revised versions of the unpolished conception.

Let us first recall the two conditions for undercutting defeat on the unpolished conception:

(i) The subject has a piece of evidence *E* that supports *p*.

(iii**) The subject has a piece of evidence *E* that destroys the support *E* provides to *p*.

The problem that afflicts the unpolished conception is, let's recall, inconsistency: (i) and (iii*) cannot simultaneously hold. In this section, I will consider a number of revisions to (i) and (iii**) and argue that none of the conceptions that ensue from those revisions works.

Here is, then, a revised version of (i):

(i2) The subject has a piece of evidence *E* that defeasibly supports *p*.

Does the attempt to characterize undercutting *defeat* in terms of (i_2) and (iii**) work? No, and the problem is obvious: such an attempt is circular, just as the attempt to characterize evidential defeat in terms of (i_2) and (ii) would be circular. Furthermore, it seems clear that there is no need to appeal to the notion of defeasibility in our characterization of evidential defeat because (i) and (ii) appear to be just fine. And, a retreat to (i) and (ii) would take us back to square one.

Here are two further ways to revise (i):

(i_3) The subject has a piece of evidence E such that that the subject does not have E' is necessary for E to support p.

(i_4) The subject has a piece of evidence E such that that the subject does not have E' is sufficient for E to support p.

However, neither of these works. Condition (i_4) is evidently unsatisfiable because that a subject does not have E' cannot suffice for E to support p: there is no piece of evidence such that a sufficient condition for that evidence to support a proposition is that a subject having that evidence not have another piece of evidence. Surely, *something else* is needed for an evidential support relation to hold. And, the problem with (i_3) is that it does not entail that E supports p; and, the joint truth of (i_3) and (iii**) leaves it open that E does not support p, in which case there is nothing to be evidentially defeated.

Furthermore, and more significantly, both (i_3) and (i_4) is beset by the same fundamental problem. Neither is true given distributional independence, viz. that evidential support relations hold independently of what pieces of evidence subjects happen to have. Different subjects might have different pieces of evidence at different times, and whether a piece of evidence supports a proposition is not a function of the contingencies that affect the distribution of evidence among subjects. If so, contra (i_3) and (i_4), that the subject does not have E' can neither be a necessary nor be a sufficient condition for E to support p. It follows that neither (i_3) nor (i_4) is true.

Now, let us turn to (iii**) and replace it by the following:

(iii_2) The subject has a piece of evidence E' that E does not support p.

It is clear that (i) and (iii_2) are jointly satisfiable. However, the problem is that if (i) and (iii_2) are true, then E' is misleading evidence regarding E: what E' says, as it were, regarding the evidential connection between E and p is false. If so, the subject has E that supports p and also E' that mistakenly says that E does not support p. Since E supports p and E' mistakenly says that E does not support p, it turns out that the subject's total evidence (viz. E and E') supports p (cf. If I win a race, and you

mistakenly say that I lose it, then I still win).[10] However, if so, (ii) is not true. So, the joint truth of (i) and (iii$_2$) guarantees that what we thereby have cannot be an example of *evidential* defeat (and *a fortiori* cannot be an example of undercutting *defeat*).

Here is another way to modify (iii**), inspired by Feldman:[11]

(iii$_3$) The subject has a piece of evidence E' that E does not support p *in this case*.

As it stands, however, (iii$_3$) is inconsistent with (i), on the plausible assumption that (i) purports to express a general truth. So, we also need to modify (i), and a straightforward way to do this is as follows:

(i$_5$) The subject has a piece of evidence E that supports p *in this case*.

It is clear that (i$_5$) and (iii$_3$) are jointly satisfiable. However, this attempt falls prey to the very same objection that afflicts the previous attempt. It is that if (i$_5$) and (iii$_3$) are true, then E' is misleading evidence regarding E in this case: what E' says, as it were, regarding the evidential connection in this case between E and p is false. If so, the subject has E that supports p in this case and also E' that mistakenly says that E does not support p in this case. Since E supports p in this case and E' mistakenly says that E does not support p in this case, it turns out that the subject's total evidence (viz. E and E') supports p in this case. However, if so, the accordingly revised version of (ii) (which reads: The subject S has a piece of evidence E' such that E and E' together do not support p *in this case*) is not true. So, the joint truth of (i$_5$) and (iii$_3$) guarantees that what we thereby have cannot be an example of *evidential* defeat (and *a fortiori* cannot be an example of undercutting *defeat*).

4. Conclusion

I have proposed and examined a number of attempts to specify the conditions for undercutting defeat and have shown that all these attempts fail. The unpolished conception suffers from internal inconsistency; the diachronic conception fails to account for some cases that must evidently count as cases of undercutting defeat, *if* there are any, and is inconsistent with distributional independence, i.e. the fact that evidential support relations are independent of what pieces of evidence subjects

[10] It is true that the subject might be *rationally* misled by E' to believe that E does not support p and thereby abandon the belief that p; but this is compatible with the point that her total evidence supports p. How can what rationality demands from a subject might come apart from what her evidence supports? For an answer, see, for instance, Christensen's (see "Higher-Order Evidence," *Philosophy and Phenomenological Research* 81 (2010): 185-215) "bracketing" account of higher-order defeat.

[11] Richard Feldman, "Respecting the Evidence," *Philosophical Perspectives* 19 (2005): 95-119, 113.

happen to have. Furthermore, the other conceptions I have examined turn out to be either circular or inconsistent with distributional independence, or the conditions they propose fail to capture what is required for evidential defeat. Absent further conceptualizations, and on the basis of these considerations, I conclude that an adequate notion of undercutting defeat is lacking.

EVERYTHING IS SELF-EVIDENT

Steven DIGGIN

ABSTRACT: Plausible probabilistic accounts of evidential support entail that every true proposition is evidence for itself. This paper defends this surprising principle against a series of recent objections from Jessica Brown. Specifically, the paper argues that: (i) explanationist accounts of evidential support convergently entail that every true proposition is self-evident, and (ii) it is often felicitous to cite a true proposition as evidence for itself, just not under that description. The paper also develops an objection involving the apparent impossibility of believing P on the evidential basis of P itself, but gives a reason not to be too worried about this objection. Establishing that every true proposition is self-evident saves probabilistic accounts of evidential support from absurdity, paves the way for a non-sceptical infallibilist theory of knowledge and has distinctive practical consequences.

KEYWORDS: evidence, reasons, basing relation

> I used to get annoyed in abstract discussions to hear men tell me: 'You think such and such a thing because you are a woman.' But I know that my only defence is to answer, 'I believe it because it is true'…
>
> BEAUVOIR[1]

We form, maintain and revise beliefs on the basis of our evidence. When we do so, the true propositions[2] which are our evidence justify our beliefs. These true

[1] Simone de Beauvoir, *Extracts from The Second Sex*, trans. Constance Borde and Sheila Malovany-Chevallier (London: Vintage Books, 2016), 6.

[2] This paper assumes a non-mentalist, factualist ontology of evidence—primarily because this is a crucial presupposition of the existing debate between Timothy Williamson and Jessica Brown. For a fairly recent overview of the landscape of the debate concerning the ontology of evidence, see Kurt Sylvan, "Epistemic Reasons I: Normativity," *Philosophy Compass* 11 (2016): 364-376. The Self-Evidence principle which I defend in this paper plays an interesting role in this debate; for instance, Bob Beddor ("Prospects for Evidentialism," in *The Routledge Handbook of the Philosophy of Evidence*, eds. Maria Lasonen-Aarnio and Clayton Littlejohn (New York: Routledge, forthcoming)) uses a version of this principle to argue against a mentalist ontology of evidence, whereas John Turri ("The Ontology of Epistemic Reasons," *Nous* 43 (2009): 490-512) appeals to considerations related to Self-Evidence on order to argue against a propositionalist (and *a fortiori*, factualist) ontology of evidence. I take no specific stance in this paper on how the Self-Evidence principle should affect the ongoing debate about the ontology of evidence; the point is just to show that, on a factualist view of evidence, true propositions are self-evident. Thanks to an anonymous

propositions are often distinct from the (possibly false) propositions for which they are evidence. For instance, one's belief that a trash bag will soon be in the basement might be evidentially justified by the true proposition that she just dropped it down the garbage chute. However, it is not obvious that a true proposition which is evidence must always be distinct from the proposition for which it is evidence. One case in which a true proposition might be evidence for itself is when its truth is directly introspectable. For instance, the true proposition *that she is in pain* might be all the evidence that an agent has for believing that she is in pain.

A more surprising suggestion is that *every* true proposition is evidence for itself:

> Self-Evidence: For any true proposition P, P is evidence that P is true.[3]

This principle says that true propositions are self-evident in an *objective* sense. A true proposition P is evidence that P in the sense of providing an objective epistemic reason to believe P, regardless of whether anyone actually *possesses* this reason as such. In other words, true propositions are self-evident without necessarily being self-evident *for* any agent. However, once an agent possesses a proposition P as evidence (for the purpose of this paper, I will take *knowing* P to be sufficient for possessing P as evidence),[4] then P will also become *subjectively* self-evident, in the sense of being an epistemic reason which this agent possesses for believing P.

The most prominent defender of Self-Evidence is Timothy Williamson.[5] Williamson sees Self-Evidence as a surprising consequence of a general probabilistic

---

reviewer on a previous version of this paper for pressing me to clarify this.

[3] There may be some exceptions to this generalisation. For instance, Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000, 188) argues that some true propositions (e.g., tautologies) cannot be evidence for anything, and therefore cannot be self-evident either. I will not be concerned with possible exceptions to Self-Evidence in this paper.

[4] See, for example, John Hyman, "How Knowledge Works," *Philosophical Quarterly* 49 (1999): 433-451; Williamson, *Knowledge and its Limits*, 184-208. Actually, following Errol Lord, *The Importance of Being Rational* (Oxford: Oxford University Press, 2018), I think that one can possess a true proposition as evidence even if she is only in a *position to know* this proposition. This makes sense of how certain truths, e.g., simple mathematical and moral truths, are subjectively self-evident (for anyone) even though people have not yet come to believe (and so, know) them. That is, since any agent (perhaps, with an adequate conceptual repertoire) is necessarily in a position to know these truths, then any agent already possesses these truths as self-evident.

Several writers have also argued that, in order to possess a true proposition P as a reason to believe another proposition Q, one must also *treat* or *conceive of* P as a reason to believe Q (see Lord, *Importance of Being Rational,* 97-124 and references therein). I will not be concerned with a potential treating condition on reason-possession in the present paper.

[5] Williamson, *Knowledge and its Limits*, 187-8; see also Timothy Williamson, "E=K, but what

approach to evidential support. On standard versions of this approach, one proposition is evidence (for some agent)[6] to believe another proposition exactly when the probability (on an agent's total background evidence) of the truth of the latter proposition conditional on the former is greater than its unconditional probability (or alternatively, as long as the conditional probability is above a certain threshold). That is, for any propositions P and Q and a subjective probability function Pr(.):

P is evidence that Q iff $Pr(Q|P) > Pr(Q)$[7]

The probability of the truth of any proposition conditional on itself is always 1 (as long as its unconditional probability is nonzero). Therefore, as a limiting case of evidential support, any true proposition which is evidence for anything provides maximal evidential support for itself. Moreover, this entailment cannot be plausibly blocked by adding conditions to the simple probabilistic account.[8] For instance, simply declaring that no proposition P is evidence for itself does not rule out the entailment that P&Q (for any arbitrary Q) is also perfect evidence that P.

For the most part, this surprising result has been greeted with an incredulous stare. It may seem deeply counterintuitive, and perhaps even circular,[9] that *any* true proposition can be evidence for itself—let alone that every true proposition is. Williamson's response is to point out that counterintuitive results are often entailed

about R?," in *The Routledge Handbook of the Philosophy of Evidence*, eds. Lasonen-Aarnio and Littlejohn.

[6] It is worth noting that a true proposition can be evidence *for a particular agent* (i.e., relative to her existing background evidence) even if she does not possess this truth as evidence. That is, the distinction between agent-neutral and agent-relative epistemic reasons does not align with the possessed/unpossessed distinction. However, since true propositions are always objectively self-evident relative to any agent's background evidence, I will not be concerned with the agent-neutral/agent-relative evidence distinction here.

[7] The unconditional probability of Q in this formula is the probability of the truth of the proposition 'prior to investigation' (Williamson, *Knowledge and its Limits*, 211). This is why, even though the probability of every true proposition conditional on itself is always 1, the unconditional probability of (almost) any non-tautological proposition is less than 1 (and therefore why conditioning this proposition on itself raises the probability of its truth).

[8] See Jessica Brown, "Evidence and Epistemic Evaluation," in *Oxford Studies in Epistemology*, Vol. 5, eds. Tamar Szabó Gendler and John Hawthorne (Oxford: Oxford University Press: 2015), 44-54.

[9] This charge of circularity, which often arises in conversation, is misguided (at least, given the present factualist ontology of evidence). Self-Evidence does not entail that all true propositions are 'self-justifying' in some objectionable sense. True propositions justify beliefs, not true propositions. However, if we understand evidence on a mentalist view (for instance, where evidence is constituted by beliefs rather than true propositions), then the circularity charge might be appropriate.

by our simplest and best logical and mathematical theories; for instance, "[c]ommon sense did not want 0 to be a number; it did not want a contradiction to entail everything; it did not want an axiom to have a one-line proof consisting of just the axiom itself."[10] Just like these cases, Self-Evidence is an instance where our best theoretical frameworks are better guides to truth than educated common sense.

Actually, it is not even clear that Self-Evidence is very counterintuitive. Many (perhaps even *most*) epistemologists accept a version of the Truth-Norm of Belief: *Believe P (if and) only if P is true*. Since valid norms are in the business of providing normative reasons, then it is natural to see the Truth-Norm of Belief as entailing that, for any proposition P, P's truth is an objective epistemic reason to believe P. In other words, the extremely intuitive principle that "truth is the aim of belief" plausibly entails that the Self-Evidence principle is valid.

However, Jessica Brown has recently emerged as a vocal opponent of Self-Evidence.[11] She has advanced two specific objections against the principle:

(1) *Non-Convergence*: Self-Evidence is a formal artefact of the probabilistic approach to evidential support, which is not entailed by other prominent approaches.[12]

(2) *Infelicity*: It is almost always infelicitous to cite a true proposition as evidence for itself; moreover, this infelicity cannot be explained away pragmatically.[13]

Both of these objections fail. Or so this paper argues. I also clarify a third possible objection to Self-Evidence involving the (im)possibility of believing P *on the basis of* the true proposition P (building on some of Brown's discussion). It is beyond the scope of this paper to resolve this final objection, but I give a reason not to be too worried about it at present.

§1 shows that the most prominent alternative approach to evidential support convergently entails that every true proposition is self-evident. §2 argues that it is often felicitous to cite a proposition as evidence for itself, just not under that description. §3 raises a potential problem involving the combination of Self-Evidence and a prominent version of the Ought Implies Can principle, and briefly explores a direction for resolving this problem. Every true proposition *really is* self-evident (including this one). §4 briefly discusses the philosophical and practical significance of this conclusion.

---

[10] Williamson, "E=K."

[11] See Jessica Brown, "Infallibilism, Evidence and Pragmatics," *Analysis* 73 (2013): 626-35; *Fallibilism, Evidence and Knowledge* (Oxford: Oxford University Press, 2018) 45-66; and "Evidence and Epistemic Evaluation".

[12] Brown, "Infallibilism," 628.

[13] Brown, "Infallibilism," 628-32; Brown, *Fallibilism*, 53-60.

1. (Non-)Convergence

Self-Evidence is a surprising entailment of the probabilistic approach to evidential support. This might be seen as a reason to accept Self-Evidence, but it can also be taken as a *reductio* of that approach. In other words, if probabilistic accounts of evidential support uniquely entail the counterintuitive Self-Evidence principle, then it might be rational to abandon these accounts in favour of their main competitors. Brown suggests that probabilistic accounts *are* unique in entailing Self-Evidence, and she thereby implies that the principle is just an implausible formal artefact of a misguided philosophical theory.[14] On the other hand, if several distinct approaches to evidential support *independently converge* in entailing the truth of Self-Evidence, then this is strong evidence in favour of the principle. This section argues that, contrary to Brown's suggestion, the most prominent competing approach to evidential support also entails that, as a limiting case, every true proposition is self-evident.

The main alternative to the probabilistic approach seeks to ground evidential support on explanatory connections between propositions.[15] This *explanationist* approach standardly emphasises the role of Inference to the Best Explanation (IBE) in scientific practice and everyday scenarios. On a basic understanding, IBE says that one true proposition P is evidence (for some agent) to believe another proposition Q as long as Q is part of Best Explanation[16] (on the agent's total background evidence) for why P is true. As Brown points out, true propositions are generally not self-explanatory; so if IBE is all there is to the explanationist approach to evidential support, then this approach does not entail Self-Evidence.

However, as a general account of evidential support, this basic version of the explanationist approach is obviously inadequate. First, the simple account cannot explain how we gain justified beliefs about simple logical consequences of known propositions; for instance, forming a justified belief that there are four animals in

[14] Brown, "Infallibilism," 628.

[15] See, for instance, Earl Conee and Richard Feldman, "Evidence," in *Epistemology: New Essays*, ed. Quentin Smith (Oxford: Oxford University Press: 2013); Kevin McCain, *Evidence and Epistemic Justification* (London: Routledge: 2014). Brown ("Infallibilism," 628) also mentions a Hempelian Hypothetico-Deductive account of evidential support, which analyses evidential support in terms of logical entailment from observables. The requirement that evidence be observable prevents unobservable true propositions being evidence for themselves, so this account does not entail (universal) Self-Evidence. However, the observable/unobservable distinction is notoriously problematic and, in any case, Hypothetico-Deductivism is no longer a prominent approach to evidential support.

[16] That is, the explanation which scores highest (and perhaps also "high enough") with respect to the explanatory virtues, i.e., simplicity, coherence, unification, etc.

one's garden on the basis of the evidence that there are two squirrels and two birds.[17] Second, the account cannot explain how we can have justified beliefs about many events in the near future, such as whether the golf ball which an agent just putted will go into the hole.[18]

There are several improved explanationist accounts available, the most prominent of which has recently been developed by Kevin McCain.[19] This account develops the basic IBE version of the explanationist approach in two separate ways. First, McCain argues that a proposition P can be evidence for another proposition Q as long as Q is a logical entailment of P (and the agent for whom it is evidence is appropriately sensitive, in some sense, to this logical connection).[20] Second, he argues that, just as we make inferences backwards along an explanatory chain via IBE, we can also make inferences *forwards* along an explanatory chain.[21] For instance, an agent can infer from the fact that she putted the golf ball in a certain way that it will go into the hole, since it goes into the hole because of how she putted it. Call this, Inference to the Best Explanatory Consequence (IBEC). This development of the basic explanationist approach says that a true proposition P can be evidence (for some agent) to believe another proposition Q as long as P would give a Better Explanation (on her total background evidence) of the truth of Q than of its falsity.[22]

Since every true proposition obviously logically entails itself, the first of these developments is sufficient to ensure that Self-Evidence is true. However, this simple argument does not give a satisfying explanation for why Self-Evidence is entailed by the explanationist approach, nor does it show that the principle obtains specifically as a *limiting case* of evidential support (as on the probabilistic approach). Therefore, there is scope for seeing this as a mere coincidence rather than as genuine

[17] See, for instance, Alvin Goldman, "Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Troubles, Reliabilism's Rescue Package," in *Evidentialism and Its Discontents*, ed. Trent Dougherty (Oxford: Oxford University Press: 2011).

[18] See T. Ryan Byerly, "Explanationism and Justified Beliefs about the Future," *Erkenntnis* 78 (2013): 229-243.

[19] See especially McCain, *Evidence and Epistemic Justification;* Kevin McCain, "Explanationism: Defended on All Sides," *Logos & Episteme* 6 (2015): 333-349. McCain's explanationist account is mentalist, in the sense that he takes an agent's evidence to be constituted by her non-factive mental states rather than true propositions. However, it is straightforward to construct a factualist version of McCain's account, which remains true to many of the motivations behind his proposal.

[20] McCain, *Evidence and Epistemic Justification*, 64-8.

[21] Compare, "upwards" and "downwards" inferences in Nevin Climenhaga, "Evidence and Inductive Inference," in *The Routledge Handbook of the Philosophy of Evidence*, eds. Lasonen-Aarnio and Littlejohn.

[22] McCain, "Explanationism," 339.

convergence. However, the second development of the basic explanationist account plugs this gap, by specifically illustrating why the Self-Evidence principle obtains as a limiting case of evidential support.

Once we recognise that we can infer both backwards and forwards along an explanatory chain, then we can combine the two sorts of inference to undergird a single evidential support relation. For instance, on the background assumption that there is a causal connection between parent/child smoking behaviour, an agent could use IBE to infer from the fact that X smokes to the proposition that X's father smoked, and then infer via IBEC to the conclusion that X's siblings also smoke.[23] In other words, it is standardly accepted that one true proposition can be evidence for another proposition when they have a *common explanation* (and the agent for whom it is evidence is appropriately sensitive to this explanatory connection). However, it is trivially true that every true proposition has a common explanation *with itself*. Therefore, the very same reasoning which showed that the fact that X smokes is evidence that X's siblings smoke also establishes that the true proposition that X smokes is evidence that X smokes.

Although this shows that true propositions are evidence for themselves, it is not immediately clear how *strong* this evidential support relation is. For instance, in the example above, the proposition that X smokes is not (much) stronger evidence for itself than it is for the proposition that X's siblings smoke. In other words, on this version of the explanationist approach, the strength of any proposition P as evidence for itself is determined by the 'goodness' of the explanatory connection (on some agent's background evidence) between P and some other proposition Q. This looks like a strange result, since by choosing a proposition Q which gives an arbitrarily good (or bad) explanation of P, it seems that P can be arbitrarily strong (or weak) evidence for itself. However, by looking at the deeper motivation behind the explanationist approach, we can see why the limiting case of this procedure entails that every true proposition *maximally* evidentially supports itself, just like on the probabilistic approach.

Recall that Brown's initial suggestion was that, on explanationist accounts of evidential support, true propositions cannot be evidence for themselves since they do not explain themselves. In other words, if explanationist accounts seek to ground evidential connections directly on the *explanation* relation, then because explanation is not a reflexive relation, Self-Evidence cannot be true. However, this is a mischaracterisation of the explanationist approach. The basic idea is *not* that any true proposition which is evidence for another proposition must explain why the latter is true, but rather the weaker claim that there must be an *explanatory*

[23] Compare, "sideways" inferences in Climenhaga, "Evidence and Inductive."

connection between the two propositions. Moreover, although *explanation* is not a reflexive relation, *explanatory connectedness* plausibly is. That is, since explanatory connections run both backwards and forwards, then every true proposition which explains or is explained by anything is necessarily explanatorily connected to itself.[24] This can be known *a priori* (and agents are plausibly already implicitly aware of it). Therefore, on any agent's background evidence, there is a guaranteed explanatory connection between any true proposition and itself, so any true proposition (which the agent possesses as evidence) provides maximal evidential support for itself. In sum, just like the probabilistic approach, adequate versions of the explanationist approach to evidential support convergently entail that, as a limiting case, every true proposition is self-evident.

2. (In)Felicity

Notwithstanding this convergence, Brown thinks that she has a knockdown objection against Self-Evidence. This is simply the observation that, except for certain special cases,[25] it is always infelicitous to cite a proposition as evidence for itself. For instance, she writes that:

> …if Morse is asked for his evidence that Burglar Bill was in the vicinity of the Central Jewellery Store at the time of the heist, it is infelicitous for him to reply by saying 'Burglar Bill was in the vicinity of the Central Jewellery Store at the time of the heist.'[26]

Moreover, Brown argues extensively that this infelicity cannot be explained away by appeal to Gricean maxims, conversational norms or other pragmatic phenomena.[27] Therefore, the infelicity of citing a true proposition as evidence for itself can only be explained by the falsity of the Self-Evidence principle.

It is not necessary to challenge these specific arguments here, except to note that this general style of argument tends not to be very convincing for those who

[24] More formally, the reflexivity of the explanatory connectedness relation is entailed by the fact that this relation is symmetric and transitive, combined with the platitude that every true proposition has some explanation (or explains something). This is not to say that the *evidential support* relation is symmetric and transitive, since explanatory connectedness is necessary but not sufficient for evidential support on the explanationist approach. However, as I make clear in the main text, evidential support *is* reflexive, since the other jointly-sufficient conditions are met; most importantly, that the explanatory connection be the 'best' on the agent's background evidence.

[25] These special cases include "self-verifying propositions, propositions concerning the nature of one's experiences, obvious logical truths, simple analytic truths, and so on" (Brown, *Fallibilism*, 51).

[26] Brown, *Fallibilism*, 51.

[27] Brown, "Infallibilism"; Brown, *Fallibilism*, 53-61.

think that pragmatic explanations come very cheap.[28] The more interesting point is that, even if Brown is correct that it is infelicitous for Morse to reassert a proposition as evidence for itself in the example above, this observation is argumentatively irrelevant.

In the example above (and in Brown's other examples), Morse is asked the *specific* question, 'What is your evidence for P?'. However, as Brown acknowledges,[29] it is dialectically open whether philosophers and non-philosophers alike are generally theoretically mistaken about the nature of evidence. That is, the proponent of Self-Evidence can accept that it is infelicitous to respond to a demand for evidence for P (under that description) by merely reasserting P, but simply attribute this infelicity to a widespread misunderstanding about the nature of evidential support. In other words, citing a true proposition as evidence for itself (under that description) is like making other assertions which, although perfectly true, contravene common sense. For instance, compare the logician who infelicitously tells a naïve audience of non-philosophers that a contradiction entails everything.

Brown argues that this impasse must be resolved in favour of her explanation of the infelicity (i.e., that Self-Evidence is false) because this remains closer to the folk conception of evidential support—and moreover, the proponent of Self-Evidence can offer no *independent* evidence in favour of the alternative interpretation.[30] This would be a weak argument even if the premises were true. However, it turns out that there *is* a strong piece of independent evidence in favour of the widespread-error interpretation of Brown's observed infelicity. It is actually often felicitous to cite a true proposition as evidence for itself, *as long as it is not cited under that description*.

Demands for an agent's evidence in shared reasoning allow us to answer two kinds of question. First, there is the explanatory question of why this agent actually believes a particular proposition P, i.e., which evidence or apparent evidence *rationalises* her belief. In other words, what are the *reasons for which* she believes P. Second, there is the normative question of why the agent *should* believe P. That is, we are interested in what (possessed) reasons *favour* believing P, even if these reasons are not the ones which motivated the agent to believe P in the first place. It is important to keep these two kinds of question distinct, since §3 of this paper discusses a potential problem for the true proposition P being the reason *for which*

[28] See, for instance, Jonathan Kvanvig, "*Fallibilism: Evidence and Knowledge*, by Jessica Brown." *Mind* 128 (2019), 1394.

[29] Brown, *Fallibilism*, 62.

[30] Brown, *Fallibilism*, 62-3.

one believes P. Therefore, although I do think that one can cite a proposition P as evidence for itself in response to both the explanatory and normative questions, I will only discuss the normative question in this section.

The important point is just that, when we ask agents for their evidence for P *by means of* asking them what reasons favour believing P, they can felicitously cite P as evidence for itself. For instance, consider the following simple dialogue:

> MURDERER'S WIFE: Just give me one good reason to believe that my husband is the murderer!
>
> HOLMES: Well, it's true! He is the murderer.

Obviously, Holmes is being uncooperative here. I think that his response is nevertheless a *correct* answer to this woman's question. It may also be *apt*, given the plausible background assumption that the wife will likely not accept alternative evidence which Homes could cite, for instance, the collection of unusual premises from which he made his brilliant and unconventional abduction. Therefore, although Holmes is not citing evidence which could rationally convince this person to change her mind, this does not mean that he is not giving a legitimate reason for believing. After all, it would be absurd for the murderer's wife to respond by saying, 'That's no reason at all to believe he's the murderer!.'

In fact, it seems (to me, at least) that, whenever one knows a particular proposition, one can correctly respond to a demand for one's reason for believing that proposition by simply (re)emphasising that it is true, although the uncooperativeness of this response would make it infelicitous in most conversational contexts. Nevertheless, there are a number of contexts in which, although one may also possess other evidence for believing a proposition P, it is *only* appropriate to cite P as a reason to believe itself.[31] For instance, there are plausibly cases of relatively unsophisticated agents who know a proposition P without having access to the grounds or method which they use for believing P. In particular, Srinivasan's recent portrait of an agent Nour,[32] who is sensitive to the fact that a particular piece of behaviour was racist without having access to what exactly about the behaviour was racist, plausibly fits this bill. If someone demanded that Nour give a reason why they should believe that this behaviour was racist, it would be incorrect for her to say that she is not aware of any reason at all. Instead, what she should say (and what it

[31] However, there are also plausibly many other contexts where it is felicitous to cite multiple reasons for believing a proposition, *one of which* is that proposition's truth. For instance: "Why should anyone think that climate change is real?" "Because *it is* real – and also look at these scientific studies, etc."

[32] Amia Srinivasan, "Radical Externalism," *The Philosophical Review* 129 (2020): 395-431.

is felicitous for her to say), is that she *does* have a reason to believe that the behaviour was racist, namely the fact that (as she just knows) *it was racist*.

The overall lesson is that it can be felicitous to cite a true proposition P (or the proposition that P is true) as a reason to believe P. One might concede this point, but reject the thesis that all reasons to believe P are evidence for P. This thesis *is* plausibly false, since there can be practical reasons for belief; but there is no sense in which P's truth could be seen as a merely practical reason to believe P. Perhaps there is also a potential verbal dispute here; but an arbitrary restriction on the extension of the concept, 'evidence,' such that true propositions cannot be self-evident even though they can be epistemic reasons to believe themselves, would obviously be unattractive.

Therefore, contrary to Brown's central objection to Self-Evidence, it is often felicitous to cite a proposition P as evidence for itself, just not under that description. Moreover, if the above discussion is correct, then the (apparent) infelicity of responding to a demand for evidence by reasserting the claim under question just results from a misunderstanding. I think we would do well to become accustomed to citing propositions as self-evident, even under that very description.

3. Basing

Although she does not rely upon it in her case against Self-Evidence, Brown's discussion hints at a final objection. She tentatively proposes a positive account of evidential support which is supposed to explain why true propositions cannot be evidence for themselves.[33] In brief, the suggestion is that one true proposition Q can be evidence for another proposition P only if it is possible for some agent to gain "first-time justification" for newly believing P on the basis of her evidence Q. Brown's contention is that it is impossible for an agent to gain "first-time justification" for believing P on the basis of the true proposition P itself, and this is why true propositions can never be evidence for themselves. In particular, if an agent tried to infer P from the premise P, then this inference would either be superfluous (since she already knew or believed that P) or circular (since she had no independent basis to believe P in the first place).[34] In neither of these cases would the agent gain first-time justification for believing P. Therefore, if the possibility of first-time justification is a necessary condition on what it takes for one proposition to be evidence for another, Self-Evidence must be false.

[33] Brown, "Evidence and Epistemic Evaluation," 54-8; *Fallibilism*, 65n.11.

[34] This is also supposed to explain why propositions like P∨P or P&Q are not evidence for P (except in special cases: see Brown, "Evidence and Epistemic Evaluation," 55-6).

There is a straightforward way to further motivate and clarify this objection. A prominent version of the Ought Implies Can principle states that a true proposition can be a reason to do a particular thing only if it is possible for agents to do this thing *on the basis of that reason*.[35] For present purposes, this principle entails that the true proposition P can be an epistemic reason to believe (i.e., evidence for) P *only if* it is possible for agents to believe P on the basis of P. However, this appears to be impossible. For instance, even moving away from the simplistic inferentialist picture of the epistemic basing relation which features in Brown's brief discussion, most standard accounts of the basing relation say that part of what it is for an agent to believe P on the basis of her evidence Q is for this agent's believing Q to non-deviantly cause her to believe P.[36] Since one cannot believe P because she believes P, on pain of explanatory circularity (or having multiple beliefs with the same content), then it must be impossible to believe P on the basis of the reason P.[37] Thus, by this version of the Ought Implies Can principle, P cannot be a reason to believe itself. True propositions cannot be self-evident.

This is a powerful argument, and although one could challenge the relevant version of the Ought Implies Can principle, I think it is very plausible. It is beyond the scope of this paper to attempt to fully resolve this objection, but there is an important reason not to be too worried about it. The argument closely parallels the well-known *No Guidance* objection to the Truth-Norm of Belief.[38] This objection says (roughly) that, although valid norms must be capable of guiding the behaviour which they purport to regulate, it would be impossible for a Truth-Norm to guide belief-formation in this way, because one would already need to believe P in order to be guided by the norm. However, although many philosophers recognise the force of the No Guidance objection, it seems that most have remained committed to the validity of the Truth-Norm even in the absence of a full resolution of the guidance problem. Even so, a number of potential solutions have been offered, which appeal,

[35] See Jonathan Way and Daniel Whiting, "Reasons and Guidance," *Analytic Philosophy* 57 (2016): 214-235, and the references cited therein.

[36] See, for instance, Keith Allen Korcz, "The Causal-Doxastic Theory of the Basing Relation," *Canadian Journal of Philosophy* 30 (2000): 525-550.

[37] Kevin McCain ("Epistemic Conservatism and the Basing Relation," in *Well Founded Belief: New Essays on the Epistemic Basing Relation*, eds. J. Adam Carter and Patrick Bondy (New York: Routledge, 2020): 201-214) points out that it is possible for one's believing P *at time t* to cause her to continue to believe P at some later time, but (as he makes clear) this would not be sufficient for one to believe P on the basis of one's evidence P.

[38] See especially, Kathrin Glüer and Åsa Wikforss, "Against Content Normativity," *Mind* 118 (2009): 31-70; Kathrin Glüer and Åsa Wikforss, "Against Belief Normativity." in *The Aim of Belief*, ed. Timothy Chan (Oxford: Oxford University Press, 2013).

for instance, to the transparency of doxastic deliberation[39] or to being indirectly guided by the Truth-Norm by means of complying with derivative evidentialist norms.[40] I think it is reasonable to suppose that the Self-Evidence principle and the Truth-Norm of Belief stand and fall together, so that if one of these strategies can successfully show that it is possible to be guided by the Truth-Norm of Belief, then it is plausible that this strategy can also show that it is possible to believe P (perhaps indirectly) on the basis of the reason P.

More specifically, a number of philosophers have argued that it *is* possible to believe a proposition P directly on the basis of the true proposition P, for instance, in successful perception.[41] Although it remains conjectural, I suspect that a similar story could be told about inference, such that when one successfully infers a proposition P from her evidence Q (and thereby comes to know P), she believes P on the basis of the true proposition P *by means* of believing P on the basis of Q.

For now, all I will say in support of this conjecture is that, (returning to the discussion in §2) just as people sometimes offer the true proposition P (or the *truth of* this proposition) as an epistemic reason to believe P, it seems that they also sometimes cite P as the reason *for which* they believe P. For instance, this is how I read the passage from *The Second Sex* which is the epigraph to this paper. In juxtaposing (but also reconciling) the causal effect of being a woman on her beliefs with the rational responsiveness of these beliefs to the objective truth, Beauvoir is saying that the truth of what she believes is not (just) a cause of her belief, but rather the *reason for which* she believes. More generally, and in contrast to some philosophers,[42] I see nothing absurd or especially problematic in the statement, "I believe P because (i.e., for the reason that) P is true." In particular, once we move away from a traditional belief-first model of the epistemic basing relation towards a the kind of competence account which has recently developed by Errol Lord and Kurt Sylvan,[43] there is no *obvious* theoretical barrier to believing a true proposition P simply on the evidential basis of that very true proposition.

---

[39] See Nishi Shah, "How Truth Governs Belief," *The Philosophical Review* 112 (2003): 447-482; Pascal Engel, "Doxastic Correctness," in *The Aim of Belief*, ed. Chan.

[40] See especially, Ralph Wedgwood, "The Aim of Belief," *Philosophical Perspectives* 16 (2002): 267-97; Daniel Whiting, "Epistemic Worth," *Ergo* 7 (2020).

[41] See especially, Ian Schnee, "Basic Factive Perceptual Reasons," *Philosophical Studies* 173 (2016): 1103-1118; and also Jonathan Dancy, "Acting in the Light of Appearances," in *McDowell and his Critics*, eds. Cynthia Macdonald and Graham Macdonald (Oxford: Blackwell, 2006): 121-134.

[42] See, for instance, John McDowell, "Response to Dancy," in *McDowell and his Critics*, eds. Macdonald and Macdonald, 134.

[43] Lord, *Importance of Being Rational*, 127-148; Errol Lord and Kurt Sylvan, "Prime Time (for the Basing Relation)," in *Well Founded Belief,* eds. Carter and Bondy, 141-174.

However, I leave it to another occasion to make sense of how it could be possible to believe a proposition P on the basis of the true proposition P itself. For now, the point of this discussion has primarily been to clarify what still needs to be done in order to defuse the Basing objection to the Self-Evidence principle (and also plausibly the No Guidance objection to the Truth-Norm of belief).

4. Conclusion

The Self-Evidence principle is a surprising convergent entailment of the prominent probabilistic and explanationist approaches to evidential support. Although it may be counterintuitive, the principle rationalises the practice of defending our claims by uncooperatively reasserting and reemphasising their truth. Even though it is unclear whether it is possible to believe a true proposition on the basis of this very true proposition, there are general reasons not to be too worried about potential problems arising from this. In particular, the popular and plausible Truth-Norm of Belief can be seen as a companion-in-guilt in this respect.

Self-Evidence is worth defending. On the one hand, since the principle is entailed by practically any probabilistic approach to evidential support, rescuing it from serious objections also saves the probabilistic approach from the threat of *reductio ad absurdum*. On the other, Self-Evidence provides the foundation for an attractive non-sceptical (and indeed, anti-sceptical) Williamsonian infallibilism about knowledge, where knowing a proposition entails that one's evidence guarantees that this proposition is true.[44]

Finally, although the principle may seem like a mere philosophical curiosity, it also has practical relevance for the prospect of doing epistemology in non-ideal political and social contexts. Some evidence (e.g., true propositions as evidence for themselves) can be rationally relied upon by individual agents and groups even though the evidence can never be cooperatively offered in a public exchange with those who disagree with the agent or group. Thus, there are dim prospects for a philosophical ideal of resolving deep disagreements by means of mere reasoning. However, there is also a positive upshot, which closely mirrors Amia Srinivasan's recent observations about the potential radical political significance of epistemological externalism in general.[45] Knowers can rationally retain their knowledgeable beliefs *just because they are true*, even when these agents are surrounded by gaslighting and immersed in bad ideology.[46]

---

[44] See Brown, *Fallibilism*, 3-9.

[45] Srinivasan, "Radical Externalism."

[46] Many thanks to Tim Williamson for extensive feedback on this paper throughout its development. Thanks also to Al Prescott-Couch and a number of anonymous reviewers.

EVIDENTIALISTS' INTERNALIST ARGUMENT FOR PRAGMATISM

Tsung-Hsing HO

ABSTRACT: A popular evidentialist argument against pragmatism is based on reason internalism: the view that a normative reason for one to φ must be able to guide one in normative deliberation whether to φ. In the case of belief, this argument maintains that, when deliberating whether to believe p, one must deliberate whether p is true. Since pragmatic considerations cannot weigh in our deliberation whether p, the argument concludes that pragmatism is false. I argue that evidentialists fail to recognize that the question whether to φ is essentially the question whether one should φ. Furthermore, the question of whether one should believe p can be answered on pragmatic grounds. The internalist argument turns out to favor pragmatism.

KEYWORDS: evidentialism, pragmatism, reason for belief, transparency, reason internalism

I. Shah's Internalist Argument for Evidentialism

Nishi Shah puts forward an influential argument for evidentialism, the view that only evidence can be normative reason for beliefs. In contrast, pragmatism maintains that pragmatic considerations can also be normative reasons for beliefs. Shah's argument is based on two theses:

> *Internalism*. R is a (normative) reason for S to φ only if R is capable of disposing S to φ in the way characteristic of R's functioning as a premise in deliberation whether to φ;[1]

> *Transparency*. The question whether to believe that p inevitably gives way to the question whether p is true. Differently put, the only way for us to answer the former question is by answering the latter.[2]

Internalism expresses the idea of reason internalism: normative reasons should be able to guide us. Notice that Internalism doesn't require that R actually

Acknowledgements: The paper is funded by Ministry of Science and Technology, Taiwan (107-2410-H-194-091-MY3). I would like to thank Hsi-Heng Cheng, Chi-Chun Chiu, and Fei-Ting Chen for their feedbacks on earlier versions of this paper.

[1] Nishi Shah, "A New Argument for Evidentialism," *Philosophical Quarterly* 56, 225 (2006): 485.

[2] Shah, "A New Argument," 481-82.

enter into the agent's deliberation whether to φ. It requires only that R *could* enter into her reasoning. Internalism looks plausible.

Transparency is a putative, psychological phenomenon that, when deliberating whether to believe p, we feel compelled to deliberate whether p is true. Only an answer to the latter question can settle the former.

Shah then argues that Internalism and Transparency together refute pragmatism. Transparency shows that the question whether to believe p can be settled only by answering the question whether p. Then, according to Internalism, only considerations that function as premises in deliberation whether p can be reasons for believing p. Since pragmatic considerations are irrelevant to the truth of p, they cannot be reasons for believing p. Pragmatism is false.

Shah's internalist argument has received considerable attention and objections. In general, the objections either argue that pragmatism is compatible with Internalism and Transparency,[3] or argue that Transparency is false.[4] I also argue that Transparency is false, but—unlike the latter whose arguments are mainly by way of counterexamples—I will offer a simple, theoretical argument for its falsity because the question whether to believe p does not entail the question whether p. Instead, what it entails is the *question whether one should believe p*. This in turn sheds a better light on Internalism. Once Internalism is properly understood, we can easily see that pragmatic considerations can satisfy Internalism because pragmatic considerations can weigh in our deliberation of whether we should believe p. Finally, I will extend my criticism to another version of the internalist argument more commonly found in the literature, so my thesis should interest a broader audience.

II. A Simple Mistake in Shah's Internalist Argument

Internalism says that reasons for believing p must be able to function as premises in deliberation whether to believe p, and Transparency says that the question whether to believe p entails the question whether p is true. Let's first examine how Transparency works. Shah offers the following explanation:

> In the sense [of Transparency] I have in mind, deliberating whether to believe that

[3] See Anthony Robert Booth, "A New Argument for Pragmatism?" *Philosophia* 36, 2 (2008): 227-231; Conor McHugh, "Normativism and Doxastic Deliberation," *Analytic Philosophy* 54, 4 (2013): 447-465; Asbjørn Steglich-Petersen, "Does Doxastic Transparency Support Evidentialism?" *Dialectica* 62, 4 (2008): 541-547.

[4] See Conor McHugh, "The Illusion of Exclusivity," *European Journal of Philosophy* 23, 4 (2015): 1117-1136; Nathaniel P. Sharadin, "Nothing but the Evidential Considerations?" *Australasian Journal of Philosophy* 94, 2 (2015): 343-361.

> p entails intending to arrive at belief as to whether p. If my answering a question is going to count as deliberating whether to believe that p, then I must intend to arrive at belief as to whether p just by answering that question. I can arrive at the belief just by answering the question whether p; however, I cannot arrive at the belief just by answering the question whether it is in my interest to hold it.[5]

According to Shah, the procedure of Transparency is like this: (1) my deliberation of whether to believe p entails my intention to arrive at beliefs as to whether p; and (2) I can arrive at the belief as to whether p only by answering whether p. (2) seems correct, but (1) is false. The reason is that I may intend *not* to arrive at beliefs as to whether p at all. To see this, let's first see how Shah argues for (1).

Shah maintains that it is a conceptual truth that the question whether to believe p entails the question whether p.[6] This is false, however. The question "whether to believe p?" is an infinitival embedded question. As Rajesh Bhatt argues, "all infinitival questions involve modality."[7] To use an example from Bhatt, "Hafdis knows where to fly" can be paraphrased into "Hafdis knows where she can/may fly." The modality involved can be deontic, in the sense that Hafdis knows where she is permitted to fly; or circumstantial, in the sense that Hafdis knows where it is possible to fly. More importantly for this context, the question "whether to φ" typically involves *deontic modality*. It is normally paraphrased into "whether S ought to φ" or "whether S should φ."[8] As a conceptual truth, therefore, the question whether to believe p *is equivalent to* the question whether S should believe p.

Some may object that, in some contexts, the question whether to φ may involve other kinds of modality. This could be true, but bear in mind that the issue here is about the *normative reason* for belief. The question we are trying to answer when deliberating about normative reasons is, without a doubt, normative. Moreover, philosophers working on normative reason often debate whether reason is *explanation of a normative fact* or *evidence for a normative judgment*.[9] In either way, we deliberate about normative reasons in order to answer normative questions.

---

[5] Shah, "A New Argument," 482.

[6] Shah, "A New Argument," 490.

[7] Rajesh Bhatt, *Covert Modality in Non-Finite Contexts* (London: Mouton de Gruyter, 2006), 117.

[8] Bhatt, *Covert Modality*, 122-23

[9] For the explanation view, see Maria Alvarez, *Kinds of Reasons: An Essay in the Philosophy of Action* (Oxford: Oxford University Press, 2010) and John Broome, *Rationality through Reasoning* (Chichester: Wiley-Blackwell, 2013). For the evidence view, see Stephen Kearns and Daniel Star, "Reasons: Explanations or Evidence?" *Ethics* 119, 1 (2008): 31-56 and Daniel Whiting, "Right in Some Respects: Reasons as Evidence," *Philosophical Studies* 175, 9 (2017): 2191-2208.

In the context of the ethics of belief, therefore, the modality involved in the question "whether to believe p" must be interpreted as deontic.

Contrary to Shah, therefore, Transparency is false. One reason why Shah thinks transparency is true is that the question whether to believe p is settled only by answering whether p. Now, we know that the question whether to believe p is equivalent to whether I should believe p, but an answer to whether p does not uniquely settle the question whether I should believe p. For there are three possible answers to this question: I should believe p, I should believe not-p, and I should believe neither p nor not-p (I should suspend believing whether p). If I decide that I should suspend believing whether p, then there is no need to deliberate whether p. Only when I decide that I should believe whether p should I answer the question whether p. So, Transparency is false.

More importantly, pragmatic considerations—such as "is p worth consideration?" "is p interesting or significant?" or "is believing p beneficial?"—can also feature in deliberating whether one should believe p. Recall that Internalism requires only that reasons for beliefs be able to function as premises in deliberating whether to believe. So, pragmatic considerations can satisfy the internalist requirement on reason for belief.

Let me explain how pragmatic considerations can function as premises in doxastic deliberation. There are at least two ways: first, by suspending doxastic deliberation and thereby suspending any consideration of relevant evidence; second, by counteracting evidence.

Consider the first way. We don't have to consider relevant evidence when deliberating whether to believe p, because we may suspend deliberation and we can do so on pragmatic grounds. For example, suppose on a stormy night Jane happens to think of the question, "Should I believe that the airport is closed?" This question can be answered by evidence. But it can also be settled by pragmatic considerations, for example, the fact that the closure of the airport does not matter to her. She can then decide that she should stop considering that question and should not seek evidence about that. Pragmatic considerations can thus answer the question whether to believe p, even in total disregard of any evidential consideration.

The second way indicates that, even when one judges that one should believe whether p, that deliberation is not settled only by answering whether p. Sometimes, even when the evidence for p is sufficient, people could still maintain that they should not believe p. To take a common example in favor of pragmatism, David's doctor tells him that he has a deadly brain tumor and is likely to die in three months without any treatment. An operation to remove the tumor would cure him, but its success rate is merely twenty percent. The failure of the operation will cost his life.

Understanding that the odds are against him, David may judge that he should believe that the operation will be successful, on the pragmatic grounds that his belief can make him less distressed and braver to face the operation.

Surely, evidentialists would want to reject my claim that David's pragmatic considerations can support the judgment that he should hold that belief. But where is their argument? To be clear, my goal here is not to refute all sorts of arguments in favor of evidentialism. My target is only the internalist argument against pragmatism. Shah argues against pragmatism by the thesis that pragmatic considerations cannot function as premises in deliberation *whether p*. This crucial thesis is based on Transparency. Transparency, however, is false. Internalism requires only that reasons for beliefs function as premises in deliberation *whether S should believe p*. Without Transparency, Shah offers no argument why pragmatic considerations cannot satisfy Internalism. Shah's internalist argument, therefore, fails to reject pragmatism.

III. The Mistake Generalized

To be clear, I am not claiming that people could *actually form beliefs* simply on pragmatic grounds. Instead, my claim is merely that people can answer on pragmatic considerations the question whether they should take any doxastic attitudes. In David's case, he probably cannot believe that the operation will be successful simply for pragmatic considerations, but he can judge that he should believe that.

Nevertheless, the inability to believe on pragmatic grounds leads to another version of the internalist argument for evidentialism, which is more commonly found in the literature.[10] It replaces Shah's internalist constraint with the following:

> *Internalism**. R is a reason for S to φ only if R is a consideration from which S could deliberate to φ.

Similarly, Internalism* does not require R to be actually the consideration from which S φs, but only that S *could* deliberate from R alone to φ. The difference between Internalism and Internalism* is this: pragmatic considerations can satisfy Internalism because they can function as premises in deliberation whether one should believe, whereas they do not satisfy Internalism* because, intuitively, we cannot form beliefs from pragmatic considerations alone. In other words, if we can

[10] See Thomas Kelly, "The Rationality of Belief and Some Other Propositional Attitudes," *Philosophical Studies* 110, 2 (2002): 163-196; Derek Parfit, *On What Matters* (Oxford: Oxford University Press, 2011); Joseph Raz, *From Normativity to Responsibility* (Oxford: Oxford University Press, 2011); Jonathan Way, "Two Arguments for Evidentialism," *Philosophical Quarterly* 66, 265 (2016): 805-818.

deliberate from a consideration to judge that we should believe p without being able to believe p, the consideration can satisfy Internalism but not Internalism*. If Internalism* is correct, pragmatism is in trouble.

Recently, Internalism* has been subject to several pragmatist objections by Stephanie Leary and Susanna Rinard, which I deem successful.[11] They try to meet Internalism* head-on, namely, by arguing how pragmatic considerations can be motivating reasons for belief. For example, Leary shows how pragmatic considerations can satisfy Internalism* by proposing the following account of motivating reason:

> *Motivating*. R is a motivating reason for which S φ-ed if and only if (i) S conceives of R as a normative reason to φ in some way; (ii) (i) disposes S to φ; and (iii) (ii) causes S to φ (in the right way).[12]

Pragmatic considerations for beliefs can satisfy Motivating because, as Leary points out, they can cause people to be more responsive to evidential considerations, which then cause people to believe. Using Leary's example for illustration, suppose Mary comes to believe that she would be much happier if she believes in God. This pragmatic consideration causes her to read famous arguments for the existence of God and befriend believers who share their religious testimonies to her. It also causes her to find evidential considerations of those kinds more persuasive than she did. Mary ends up believing in God. This is how people can deliberate from pragmatic consideration to believe, so pragmatic consideration can satisfy Internalism*.

Leary's argument is convincing. By giving an account of how pragmatic considerations can lead us to believe, however, I think she concedes Internalism* too much. For it's wrong to accept an internalist constraint that demands an account of how normative reason for φ-ing moves us to φ. In other words, Internalism* should be rejected. Internalism is a better account of the internalist constraint on normative reason. For Internalism, however, (2) and (3) are unnecessary, because it requires only that pragmatic considerations can function *as premises in reasoning that leads to the judgment whether S should believe*. It doesn't matter how someone deliberates from her judgment that she should believe p *to actually believe p*.

To show why Internalism* should be rejected, we must see why philosophers accept the internalist constraint (Internalism or Internalism*). The rationale behind it is a plausible idea that normative reasons must be able to *guide us*. But guiding us

[11] See Stephanie Leary, "In Defense of Practical Reasons for Belief," *Australasian Journal of Philosophy* 95, 3 (2016), 529-542; Susanna Rinard, "Against the New Evidentialists," *Philosophical Issues* 25, 1 (2015): 208-223; Susanna Rinard, "Believing for Practical Reasons," *Noûs* 53, 4 (2019): 763-784.

[12] Leary, "In Defense," 535.

to what? Internalism requires only that reasons guide us to judge that we should φ, Certainly, our judgment that we should φ will guide us further to deliberate how we can φ. But the deliberation of how we can φ is not part of the normative deliberation of whether we should φ, because the former question is a factual question about how to make φ-ing happen. Why should we accept a more demanding requirement on guidance, such as Internalism*?

To support Internalism*, Jonathan Way explains the idea of guidance as follows:

> Reasons are supposed to guide us and the way in which reasons guide us is through reasoning. … The basic thought is normative: reasons are what should guide us, and so there must be a good route from our reasons to the responses they support. Reasons must be *premises of good reasoning*.[13]

On the face of it, Way's account of guidance is similar to Shah's: reason must function as premises in our reasoning. Besides, Way says: "reasoning is directed at a question."[14] To understand their difference, therefore, I suggest looking into the questions that Internalism and Internalism* are directed at respectively.

For Internalism, the question is *whether to φ*, or equivalently, *whether S should φ*. Hence, doxastic deliberation, for Internalism, looks like this: "P1, P2, P3…, so S should believe p." On the other hand, according to Way (in another paper on guidance coauthored with Daniel Whiting), doxastic deliberation for Internalism* proceeds like this: "when you φ for the reason that p, we can think of you as engaging in a piece of reasoning: 'p, so I'll φ'."[15] To get myself φ-ing, then, doxastic deliberation should be able to lead me to believe. Thus, Way says: "It is good reasoning to move from believing p, q, r. . . to believing c only if 'p, q, r. . . , so, c' is a good argument."[16] Since pragmatic considerations cannot bridge the gap between my judgment that I will believe p to my believing that p, Way concludes that pragmatism is false.

Now, it's clear that Way's argument is flawed. In his picture of reasoning to believe, there are two steps of deliberation:

> First, the deliberation of *whether I will believe p*;
>
> Second, the deliberation of *how I can believe p* (if I decide that I will believe p).

[13] Way, "Two Arguments," 814; my italics.
[14] Way, "Two Arguments," 816.
[15] Jonathan Way and Daniel Whiting, "Reasons and Guidance (Or, Surprise Parties and Ice Cream)," *Analytic Philosophy* 57, 3 (2016): 220.
[16] Way, "Two Arguments," 815.

The above quotation from Way focuses on the second step. However, the second step is inessential in our normative deliberation of whether to φ. So, even if only evidence can be featured in the second step, we cannot conclude that reason for belief can only be evidence.

I have two reasons for the inessentiality of the second step. First, the issue here is about normative reason so that the question at issue must be normative, but the question at the second step is not normative. Second, how the deliberation at the second step proceeds is predominantly determined by the first step, and pragmatic considerations bulk large at the first step.

First, the debate between evidentialism and pragmatism here is about whether non-evidential considerations are normative reasons for beliefs. As I've discussed, normative reason is usually considered explanation of normative fact or evidence for normative truth. Hence, when we deliberate normative reason, we are aiming at answering a normative question. This is the first step in Way's picture.[17] Once I make the normative judgment as to what I should or will φ, naturally I deliberate the second step concerning how I can achieve φ-ing. The second step, however, is not normative. It's merely a factual question about how I can make φ-ing happen. Therefore, the reasons function as premises at the second step are not normative because they are not functioning as explanation or evidence for normative truths. Therefore, even if only evidence can feature in the deliberation of how I come to believe p, this does not demonstrate that only evidence is normative reason for belief because that deliberation is not normative.

Second, how the first question is answered will shape how the second step is taken. As I've argued, I can decide whether I will believe whether p purely on pragmatic grounds. If I am interested in whether p and decide that I will believe whether p, naturally I will weigh evidence for and against p. On the other hand, if I am not interested in whether p and decide to suspend believing whether p, then the second step is skipped and I don't need to seek any evidence about p. Therefore, whether the second step will be undertaken can be determined entirely by pragmatic considerations. This demonstrates that the second step is not essential to the deliberation of whether to believe, so the fact that only evidence can feature in the second step does not show that pragmatism is wrong.

To illustrate my idea more concretely, let's consider an example of reason for action (note that Internalism* is not confined to belief). Imagine a spacecraft is deviating from the track of returning to Earth. The chief engineer at NASA tries to calculate the right path and speed to bring it back to the track. She reasons: "I must

[17] For the same reason, "I will φ" should be understood as expressing a demand, like "You will pay back the money you owe me," rather than predicting the future.

get the correct answer, or otherwise the astronauts die and the mission fails." She then gathers all the required data and gets the correct answer through repeated calculations. Note that there are also two steps in her deliberation to get the correct answer: first, she deliberates what the task she must undertake is; and second, how she can accomplish it. To get the correct answer, similarly, she must exclude pragmatic considerations and consider only evidence. But it would be absurd to conclude that pragmatic considerations are not reasons for getting the correct answer. For it is because the moral and pragmatic stakes are high that she judges that she must get the correct answer at the first step, which guides her to judge at the second step that she must examine evidence and only evidence to make sure the answer is correct.

Hence, the second step in Way's picture is not part of our normative deliberation, but a step demanded by the judgment of our normative deliberation. It is purely instrumental to achieve the goal of the normative judgment set at the first step.[18] That's why I think that Leary concedes to Internalism* too much. To play the role of guidance, it is sufficient for reasons to function as premises in one's deliberation of whether one should φ. An account of how normative reasons cause one to φ is rather unnecessary.

Some might object that an account of the causal relation between judging that S should φ and S's φ-ing is not superfluous because it could be the case that S cannot φ. However, if one accepts an internalist constraint, one should also accept the principle that reason (ought) implies can.[19] So, if it is true that S cannot φ, then there is no reason to φ for S. Therefore, an account of how a consideration in favor of φ-ing disposes one to φ is not essential to the question whether it is a reason to φ.

To generalize, when articulating the internalist argument, evidentialists share the same mistake: they all forget that the first step when deliberating about normative reasons is to deliver the normative judgment concerning whether one should φ. Somehow, they all skip that normative judgment and jump from deliberating reasons to φ *to φ-ing*. Why is that the case? I could only venture to speculate that they seem to be misled by the phrase "reason *to φ*" or "whether *to φ*" into thinking that the endpoint of deliberating reasons is *φ-ing per se*. But it is wrong. As Shah and Way both maintain, reasoning is directed at a question and reasons function as premises in the reasoning to arrive at an answer. And the answer given by normative reason is a normative judgment that one should φ.

---

[18] In her "Believing for Practical Reasons," Rinard also argues that evidence is merely a means to believing, so how one comes to believe is not essential to an account of why pragmatic considerations are normative as well as motivating reasons for beliefs.

[19] See Way and Whiting, "Reasons and Guidance."

In conclusion, the internalist constraint on reason does not support evidentialism. Reason internalism plausibly requires that reasons should guide us. And reasons guide us through functioning as premises in our deliberation of normative questions. Unfortunately, evidentialists jump from deliberating whether one should φ to φ-ing, ignoring the task of normative deliberation is to answer the question of whether one should φ. Moreover, one can decide whether one should φ solely on pragmatic grounds. True, if one decides that one should believe whether p, one should answer whether p presumably by seeking evidence. But this is irrelevant. The internalist constraint correctly maintains that the function of normative reason is first and foremost to answer the normative question whether one should φ. Pragmatic considerations can satisfy the internalist constraint.

A VIRTUE THEORETIC ETHICS OF INTELLECTUAL AGENCY

Shane RYAN

ABSTRACT: There is a well-established literature on the ethics of belief. Our beliefs, however, are just one aspect of our intellectual lives with which epistemology should be concerned. I make the case that epistemologists should be concerned with an ethics of intellectual agency rather than the narrower category of ethics of belief. Various species of normativity, epistemic, moral, and so on, that may be relevant to the ethics of belief are laid out. An account adapted from virtue ethics for an ethics that goes beyond the ethics of belief is defended. The main claim advanced here is that we should act as the virtuous agent would characteristically act in the circumstances. This claim is supported with reference to a number of examples, as well as considerations informing virtue ethics. An acknowledged feature of this account is that it provides limited guidance regarding right action in intellectual agency. While the account draws on virtue responsibilism to offer guidance, the case is made that it's a mistake to think that an account in this area can provide a successful decision procedure.

KEYWORDS: virtue theory, ethics of belief, responsibilism

1. Belief or Intellectual Performance?

Ethics of belief is a recognised area within epistemology. This section makes the case that epistemology should be concerned with the ethics of intellectual agency rather than the overly narrow ethics of belief. The position is defended by observing that the ethics of belief is too narrow to capture various aspects of our intellectual lives which are deserving of the attention of epistemology. On the other hand, the ethics of intellectual agency encompasses these various aspects of our intellectual lives, many of which are exciting new topics of study in epistemology. Having made the case that epistemology should be concerned with an ethics of intellectual agency, in the sections that follow I defend a virtue theoretic account of the ethics of intellectual agency.

The ethics of belief debate is concerned with what, if any, norms govern "our habits of belief-formation, belief-maintenance, and belief-relinquishment."[1,2]

[1] Andrew Chignell, "The Ethics of Belief," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Spring 2013 Edition).

[2] Henceforth, I'll simply write "belief-formation," rather than "belief-formation, belief-

Although the question is important, "how should we act as intellectual agents?" is a more basic normative question for epistemology. The "should" in the above question should be understood in a performative sense. Examples include how the ice skater should land after a leap, how the golfer should swing his club, and so on. While such performances are not simple matters of will, after all the amateur can't simply so perform, they are the result of practice and self-cultivation, just as we can talk of how an aircraft should perform, say, at a particular altitude in particular weather conditions.[3] Theorist can disagree about the extent to which intellectual performance is voluntary.

We may nevertheless wonder why there is a well developed literature on the ethics of belief-formation in which a variety of evidentialist and non-evidentialist, deontological, consequentialist and virtue epistemological positions have been defended.[4] Perhaps on a very narrow conception of epistemology, whereby epistemology is just concerned with knowledge and the most basic way we have of being in the running for knowledge is believing, the exclusive focus on belief-formation makes sense.[5] This, however, is an overly narrow vision of the concerns of epistemology. Epistemology is plausibly also concerned with epistemic goods other than knowledge, such as intellectual virtue, good judgement, understanding, and wisdom.[6]

---

maintenance, and belief-relinquishment." Unless I indicate otherwise, "belief-formation" should be taken to also cover belief-maintenance and belief relinquishment.

[3] For more on a performance approach to epistemic agency based on functioning, see Ernest Sosa, *Judgment and Agency* (Oxford: Oxford University Press, 2015).

[4] Of course this restricted focus to belief-formation may in some cases merely be the result of authors' interests, an existing tendency in the literature to restrict focus to belief-formation, and so on. Of course such explanations aren't philosophically well motivated grounds for restricting our focus to belief formation. Indeed, I haven't found an argument for a restricted focus on belief-formation rather than the broader scope or something like the broader scope being proposed here.

[5] Belief as "most basic" here also requires that there is not another epistemic standing that is equally basic. So if acceptance is equally basic but separate from belief, then the exclusive focus on belief remains unexplained.

[6] For Jonathan Kvanvig, epistemology is the study of purely theoretical cognitive success. Such success includes, but is not limited to, making sense of the course of experience, carrying out an intellectually responsible or blameless inquiry, and an empirically adequate theory (Jonathan Kvanvig, "Truth Is Not the Primary Epistemic Goal," in *Contemporary Debates in Epistemology*, eds. Matthias Steup, John Turri, and Ernest Sosa (Oxford: Blackwell, 2013), 352-353). Linda Zagzebski, writes that epistemology's subject matter goes beyond "the study of knowledge and its components," remarks that she thinks that the broadest way to characterise epistemology is as "the study of right or good ways to cognitively grasp reality" (Linda Zagzebski, *On Epistemology* (Belmont, CA: Wadsworth, 2009), 8).

The question as to how we should act as intellectual agents is broader than the how should we believe question, as it pertains to the norms that govern our intellectual agency, which includes norms governing belief-formation. In other words, the question, by asking how we should act as intellectual agents, also asks how we should form our beliefs. "Intellectual agency" should be broadly construed and is intended as a handle for various aspects of our intellectual lives not covered by the ethics of belief. It includes belief-formation, judgement-formation, acceptance-formation, seeking evidence, seeking understanding, exercising intellectual courage, being epistemically creditworthy, and so on.[7] In fact, what in particular falls into the category of intellectual agency is not very important for my purposes here. What is important is that intellectual agency includes much more than mere belief-formation. An implication of this point is that an account of the ethics of belief formation is then insufficient to be an account of intellectual agency or, in other words, an account of our intellectual lives.

Perhaps one might hold that an ethics of something implies the potential for an agent being responsible for that thing, and belief-formation is the most basic act of intellectual agency for which we can be responsible. But this isn't right either. We can also judge, seek further evidence, exercise intellectual courage, and so on; none of which are reducible to belief, while each is something for which we can have an ethics.[8]

There is another way one might defend the exclusive focus on belief. It might be thought that although there might be acts of intellectual agency as basic as belief-formation, no such acts are basic to an epistemic good. The thought might be that as epistemologists we are concerned with more than knowledge, but for no other epistemic goods with which we are concerned is an act of intellectual agency other than belief-formation most basic. This, however, seems implausible. Blameless inquiry seems to be an epistemic good and perhaps one can have testimonial knowledge based on acceptance rather than belief.[9] Indeed, blameless inquiry seems

[7] Intellectual agency is also clearly related to the growing work on the topic of epistemic paternalism. For relevant work, see Michel Croce, "Epistemic Paternalism and the Service Conception of Epistemic Authority," *Metaphilosophy* 49, 3 (2018): 305-327 and Shane Ryan, "Paternalism: An Analysis," *Utilitas* 28, 2 (2016): 123-135.

[8] Of course we can expect that the quality of one's judgement and seeking of further evidence will partly depend on one's beliefs. It's simply that their quality depends on more than the quality of my beliefs and that therefore an ethics solely of belief would be overly narrow.

[9] See Hookway for a defence of the view that the primary concern of epistemology is how we can engage in inquiry, deliberation, and other such activities, and that knowledge and justification are subordinate concerns (Christopher Hookway, "Epistemology and Inquiry: The Primacy of Practice," in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press,

to be precisely the sort of thing that would be part of the subject matter of an ethics of intellectual agency.[10]

An ethics exclusively focused on belief provides limited guidance for our intellectual lives. Any ethics of belief will say that in various circumstances we shouldn't form a belief or that we should only form a very weak belief. But what then? We may be in a situation in which a judgement has to be made one way or another.[11] There may be various epistemic factors bearing on such a judgement. Clearly, faced with such a judgement, our intellectual agency can go beyond believing and not believing. There may be many things we should think, attend to, doubt, accept, investigate further, and so on. An ethics that just deals with belief is an impoverished ethics of intellectual agency and is out of step with recent trends in epistemology. An ethics of intellectual agency is an ethics for our epistemic lives, an ethics for the sort of epistemological situations that we face.[12] An ethics exclusively focused on belief-formation falls short of this.

In fact, while "ethics of belief" is a recognised area of study in philosophy, recognition of "ethics of intellectual agency" as encompassing ethics of belief promises to better identify for scholars closely related topics of discussion that currently risk being examined in isolation. By identifying the topics discussed above as discussions in the ethics of intellectual agency rather than some being discussions in the ethics of belief and others either falling into different demarcated areas of epistemology or not falling into any clearly demarcated area of philosophy, we also stand to better structure philosophical discussion. Of course this approach would

---

2006), 95-110).

[10] It also seems like something a virtue theoretic approach to intellectual agency would be particularly well suited to addressing.

[11] One reason to think that judging and its ethics should be accorded greater significance by epistemologists is that often what it is we believe isn't transparent to ourselves or we may simply not have a belief on the relevant matter. Do I believe that doing a PhD would be good for the student? Do I believe the politician who says that the employment situation for young people will improve significantly very soon? Is this bike shop ripping me off by charging me this amount to repair my bike? Saying how one should form one's belief in such situations isn't particularly helpful. Maybe I will believe in the right way, but if it's not clear to me what I believe on such matters or I only have a partial belief, then my believing in the right way may not be very helpful. I'll still need to act in many such cases. What should I advise the student? How should I respond to the bill with which I'm presented? How we should form judgements is plausibly an appropriate concern for epistemology. For more on the epistemological significance of judgement, see Sosa, *Judgment and Agency*.

[12] See Kvanvig for a similar point (Jonathan Kvanvig, *The Intellectual Virtues and the Life of the Mind: On the Place of the Virtues in Contemporary Epistemology* (Savage, Maryland: Rowman and Littlefield, 1992), 187).

encourage scholars to consider whether broader normative accounts might be possible that give us unified answers a host of topics that have up until now been treated separately. The account I defend in the next part of this paper suggests that this is indeed possible in a limited way, although of course plausible bolder accounts may be possible.

Upon consideration of this new area of epistemology, an important question arises as to the nature of the normativity involved. Is the type of norm governing intellectual agency only ever epistemic or might our intellectual agency, including belief-formation, sometimes, be governed by norms other than epistemic norms?[13] I take it that non-epistemic norms include moral norms, filial norms, prudential norms, and all things considered norms.[14] Consider the evidentialist answer to the question as to how should one believe, which claims that one should believe, or form one's belief, in accordance with the evidence. The argument offered in support of William Clifford's particular evidentialism is moral, while other supporters of evidentialism, such as Earl Conee and Richard Feldman, offer an epistemological ground for their evidentialism.[15] While we should want to know which type or types of norm may govern our beliefs and intellectual agency generally, we should also want to know, if it's indeed possible to know, how particular norms of various types

[13] Susanna Rinard makes the case that the question "What Should I Believe?" is answered by the broader question "What should I do?" (Susanna Rinard, "Equal Treatment for Belief," *Philosophical Studies* 176, 7 (2019): 1923-1950).

[14] Prudential norms and all things considered norms needn't have the same extension. Prudential norms are governed by whatever makes a person's live go best for that person. So, for example, if hedonism is correct, then our prudential norms will reflect that. All things considered norms, however, may differ from such prudential norms. Say, for example, that an action would produce some pleasure for an individual but would lead to severe hardship for that person's community, then the prudential norm and all things considered norm could come apart. One might insist that prudential and all things considered norms always coincide, the point though just is that conceptually they needn't. Similarly, an ethicist who holds that morality requires impartiality may still hold that we have, say, special (non-moral) duties to our friends and family. As such, she may hold that there are also filial norms, "filial" here being broadly construed. Filial norms or moral norms are relevant to our discussion if one holds that it is sometimes permissible or even required to be partial in belief; say, for example it is permissible to believe well of one's friend with respect to some matter in the face of a balance of evidence to the contrary.

[15] William Clifford, "The Ethics of Belief," in *The Ethics of Belief and Other Essays*, ed. T. Madigan (Amherst, MA: Prometheus, 1877/1999), 70–96 and Earl Conee and Richard Feldman, *Evidentialism: Essays in Epistemology* (Oxford: Clarendon Press, 2004). See also Feldman's discussion of an alternative, epistemically grounded evidentialist ethics of belief which he contrasts with Clifford's account (Richard Feldman, "The Ethics of Belief," *Philosophy and Phenomenological Research* 60, 3 (2000): 667-695).

interact in determining how we should act intellectually.[16] However, whether it's misleading to talk of different types of normativity governing intellectual agency, saying that different norms may sometimes come into conflict, should be considered.[17]

2. An Agent-Centred Ethics of Intellectual Agency

Our goal, based on the previous section, should be to provide an ethics of intellectual agency rather than a mere ethics of belief. In the rest of the paper I defend a particular account of the ethics of intellectual agency. More specifically, I defend a virtue theoretic account of the ethics of intellectual agency.[18] The view is as follows:

> An agent should act intellectually as the virtuous agent would characteristically act in the circumstances.

Before we get to the defence of this view, some relevant background in virtue theory in epistemology and ethics is provided. The prospects of a virtue theoretic ethics of intellectual agency are considered with special attention given to the distinctive issues of action guidance and decision procedure. While the case is made against adopting a particular decision procedure, the action guidance that a virtue theoretic approach can provide is outlined in this section and the next.

For those of us already working in virtue epistemology, exploring the prospects of a virtue theoretic approach seems a natural place to start in developing an ethics of intellectual agency. Work being done in virtue responsibilism looks a good basis for just such an approach.[19] A feature of virtue responsibilism is that some theorists tend to be less concerned with providing an account of knowledge, and so less concerned with good believing, and more concerned with explaining the intellectual virtues.[20] As such, this approach then seems to have the potential to

---

[16] Interestingly, one might see William James as not denying evidentialist epistemic norms, but rather holding that sometimes epistemic norms can be trumped by prudential norms (William James, "The Will to Believe," in *The Will to Believe and Other Essays in Popular Philosophy*, eds. F. Burkhardt et al. (Cambridge: MA, Harvard, 1896/1979), 291–341).

[17] I discuss these issues more later in the paper.

[18] Obviously we could accept the foregoing section while disagreeing with the particular account of intellectual agency provided here.

[19] Prominent virtue responsibilist theorists include Montmarquet, Zagzebski, Baehr. See James Montmarquet, *Epistemic Virtue and Doxastic Responsibility* (Lanham: Rowman and Littlefield. 1993); Linda Zagzebski, *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996); Jason Baehr, *The Inquiring Mind* (Oxford: Oxford University Press, 2011).

[20] Jason Baehr, "Four Varieties of Character-Based Virtue Epistemology," *Southern Journal of Philosophy* 46, 4 (2008): 469-502.

inform the broad scope of the ethics of intellectual agency—it has the resources to say more than how we should believe.

That there is this discourse on the intellectual virtues then provides a ready platform to develop a virtue theoretic ethics of intellectual agency in one regard. We can draw on existing accounts of the intellectual virtues to develop our ethics of intellectual agency. We can, for example, examine accounts of intellectual humility and open-mindedness as a basis for developing a responsibilist virtue ethics of intellectual agency. In particular, we can draw on virtue responsibilism to develop answers as to how we should judge and inquire, as well as evaluate topics such as the epistemic goals of education and the permissibility of epistemic paternalism.

On the other hand, much less work has been done with regard to providing a unified account of the intellectually virtuous agent and there is no existing broad account of the ethics of intellectual agency to build on.[21] In order to find work on which to build, I turn instead to another virtue theoretic approach, virtue ethics. Drawing on virtue ethics marks a departure from typical approaches to virtue by epistemologists; and the Aristotelian approach favoured provides, as we shall see, an elegant way of dealing with seemingly conflicting norms (epistemic, moral, and so on) regarding topics in the ethics of intellectual agency discussed at the end of the first section.

Virtue ethics is famously agent-centred rather than act-centred. This means that the proper focus of ethics according to the virtue ethicist is the agent. In order to assess the moral worth of an action we should first examine the moral character of the agent who performed that action. This is in contrast to rival normative approaches such as consequentialism and deontology, according to which we can provide a moral assessment of an action without first assessing the moral character of the agent whose act it is.

Various forms of consequentialism and deontology provide us with very clear guidance as to how to act. A classical utilitarian—a particular sort of consequentialist, for example, holds that morality requires that we act in such a way so as to bring about the greatest possible balance of happiness. A challenge put to the virtue ethicist is to explain what action guidance her normative approach provides. The thought is that a normative approach in ethics should provide action guidance and that without providing action guidance virtue ethics risks not being a stand alone normative approach, though it might beneficially supplement other normative approaches such as consequentialism or deontology. It might supplement them in so far as filling out details of the character who acts well morally.[22] Taking our cue from

[21] Sosa would, however, be a natural place to start (Sosa, *Judgment and Agency*).

[22] Rosalind Hursthouse, "Normative Virtue Ethics," in *Ethical Theory: An Anthology*, ed. Russ

critics of virtue ethics, we might also wonder how virtue ethics can provide a basis for a virtue theoretic ethics of intellectual agency if virtue ethics provides no action guidance. In other words, how can virtue ethics help us provide a virtue theoretic answer as to how we should act intellectually if it doesn't say anything about how we should act in the moral realm?

Fortunately, virtue ethicists do have a response to the action guidance challenge. Rosalind Hursthouse defends virtue ethics against the charge that it fails to provide action guidance. According to Hursthouse, an act is right if, and only if, "it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances."[23] Of course, it isn't always clear to us how a virtuous agent would act in particular circumstances. Even granting that this is the case, virtue ethics still tells us to act virtuously (honestly, with charity, courageously, and so on) and not viciously (dishonestly, with meanness, cowardly, and so on). Hursthouse acknowledges, however, that there may be situations in which it's not clear what the virtuous thing to do is.[24] It might seem that one virtue requires me to act in one way, but that another virtue requires me to act in another, conflicting way.

While having an account of intellectual agency according to which we should act as the virtuous agent would act intellectually seems like a promising basis for a virtue theoretic account of the ethics of intellectual agency, there is an analogous problem in the epistemic sphere, especially if one's sole focus is the intellectual virtues, to that of the one faced by virtue ethics. It's well and good to say I should act as the virtuous agent would act intellectually but should I be intellectually courageous and express my opinion on a controversial issue, or should I be intellectually conscientious and stay relatively muted on the issue until I have learnt more about that issue? In cases in which there is an apparent conflict between what various virtues require, one response is that we're required to use our practical wisdom (*phronesis*) to determine which virtue we should act upon. Now, however, the charge of lack of action guidance resurfaces. Telling us that we need to use our practical wisdom doesn't give us an answer as how we should act.[25] Should we

Shafer-Landau (Oxford: Blackwell Publishing, 2007), 701.

[23] *Ibid.*, 703. Zagzebski provides a very similar account of right action, which forms part of what she calls an exemplarist virtue theory. According to Zagzebski's moral theory, the example provided by the virtuous agent is "primary" or fundamental. Treating the example as primary, however, risks raising a Euthyphro-style dilemma (Linda Zagzebski, *Exemplarist Moral Theory* (Oxford: Oxford University Press, 2017); Linda Zagzebski, "Exemplarist Virtue Theory," *Metaphilosophy* 41, 1 (2010): 41-57; Linda Zagzebski, *On Epistemology*).

[24] Hursthouse, "Normative Virtue Ethics," 706.

[25] For big picture accounts of how the wise act, see Stephen Grimm, "Wisdom," *Australasian Journal of Philosophy* 93, 1 (2015): 139-154 and Shane Ryan, "Wisdom: Understanding and the

express our opinion or should we keep stumm about what we think until we're better informed?

But why think that a normative theory can give us correct guidance as to what to do in every moral or epistemic situation?[26] While it would be useful to have a decision procedure that we could apply in every situation that would correctly tell us what the morally or epistemically right thing to do is in those various situations, no such plausible decision procedure has been found.[27] Furthermore, there doesn't seem to be any reason to think that we could discover such a decision procedure.[28] If we accept this, then the criticism that virtue theory fails to be action guiding in the way being considered here is a moot point. It's not a weakness of virtue theory that it fails to be action guiding in the way described, if no normative approach can provide us such action guidance.

If one holds that no correct decision procedure is available, then a normative approach in epistemology centred on action seems problematic. Simply to accept this and leave matters there would result in an unsatisfyingly fragmentary normative approach in epistemology.[29] If, however, we think that aetiology matters in epistemology, that it bears significantly on the epistemic status of an action, then a natural locus for our theoretical focus is the source of the action, the agent. By

---

Good Life," *Acta Analytica* 31, 3 (2016): 235-251.

[26] In fact, some utilitarians and deontologists hold that they too must require that agents in some cases employ practical wisdom in order for their respective rules to be applied correctly (Rosalind Hursthouse, "Virtue Ethics," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Fall 2013b Edition)).

[27] Of course this is unlikely to satisfy theorists who believe that there is such a decision procedure. To go beyond pointing out that there's no plausible candidate in the literature and make the case against various proposed decision procedures, however, would take us too far afield. Furthermore, note that a decision procedure purports to tell us how to act morally in a situation but, depending on the precise moral theory, it may not tell us whether a particular situation calls for moral action or not. To put the point differently, such a decision procedure may not help adjudicate between, say, conflicting moral and prudential norms in a given situation; we'll get an answer as to what the moral thing to do in the particular situation is, not necessarily whether we should do the moral thing rather than the prudential thing. I'll return to this later in the paper.

[28] While this is the position of various virtue ethicists, moral particularists are natural allies on this point. According to the strongest version of moral particularism, developed by Jonathan Dancy, there are no moral principles that hold irrespective of situation (Jonathan Dancy, *Ethics without Principles* (Oxford: Clarendon Press, 2004).

[29] Such a deficiency of unity worry is reminiscent of the charge made against ethical intuitionism and is discussed in David McNaughton, "An Unconnected Heap of Duties?" *Philosophical Quarterly* 46, 185 (1996): 433.

shifting our theoretical focus back from actions to the agent, we are better placed to develop a normative approach in epistemology that enjoys theoretical unity.[30]

3. Guidance for Good Intellectual Agency

By adopting this approach for our account of the ethics of intellectual agency, we are able to provide some guidance as to how we should form our beliefs. More significantly, however, is that the basic normative approach outlined provides guidance for good intellectual agency generally. This is an important theoretical advantage. As discussed, the proper scope of concern for an epistemologist is not merely how we should believe but rather concerns our epistemic agency more generally. Having a theory that accounts for both belief and epistemic agency more generally is important in so far as it facilitates a more unified approach to the ethics of intellectual agency. A theoretical approach that only concerned doxastic responses wouldn't be helpful in cases in which non-doxastic responses are also relevant.

The cases that follow, however, underscore the challenges facing a unified theoretical approach to the ethics of intellectual agency. The subsequent discussion develops an approach that attempts to meet these challenges without rejecting anti-theory. In these cases the epistemic agents described should respond differently doxastically to the same case.[31] The first two cases show this. In the first case, epistemic norms support differing intellectual responses. In the second case, differing intellectual responses are supported by differing norms but those differing responses are both all things considered appropriate. The final case shows that sometimes an appropriate intellectual response goes beyond merely forming beliefs in certain ways.

> Case 1
>
> An eyewitness provides testimony that pertains to a crime. The eyewitness appears sincere and competent in the relevant respects, and is in fact sincere and competent in the relevant respects. The testimony is heard by both a layperson and a police investigator. It's permissible, perhaps required, of a layperson in normal circumstances to believe the testimony. For a police investigator, who is just beginning to investigate the case, intuitively it is impermissible for him to believe

[30] It should be noted, however, that the return to virtue based approaches in both the epistemic and moral domains is relatively recent and those approaches are still being developed. In epistemology, even though virtue theoretic approaches are highly influential in the area of accounts of the nature of knowledge, they are much less influential in the ethics of belief debate.
[31] This is one aspect of the approach that goes beyond the adaption of Hursthouse's virtue ethics.

the testimony, rather he should withhold belief.[32]

Case 2

Suppose a person is diagnosed with a life threatening illness. She is told that the survival rate for people with the illness is 20%. If, however, a person believes that they will survive, then the survival rate improves to 40%. In this case it is clearly permissible, perhaps even required, that the patient believe that she will survive. This is regardless of whether she forms her beliefs on the basis of, say, wishful thinking, or not. For a doctor aware of the details of the sick patient's case, the illness and survival odds and so forth, and supposing there's no significant extra evidence that the doctor has, a belief that the patient will survive is not permissible.[33]

Case 3

Suppose a child cries inconsolably. The reason for his crying is completely unclear. In such a case it is not enough for a primary care-giver to believe in accordance with the evidence as to why the child is crying inconsolably. It's appropriate, rather, to try to understand why he is crying inconsolably.[34]

The first two cases show that how one should form one's beliefs may differ depending on one's situation. In the first case, the police investigator's position is such that intuitively it is inappropriate for him to simply believe the testifier, while it is appropriate for a layperson to believe the eyewitness. The case is interesting in that *norms direct the agents' belief-formations differently on epistemic grounds*.[35]

---

[32] One might claim that it doesn't matter what the police investigator believes, it only matters what he does. But given the plausibility that believing would have a psychological impact on the police investigator and, in turn, on the investigation of the case, it's preferable that the investigator withhold belief. If the police investigator already believes that a particular person is the perpetrator of a crime, then we expect that the way he questions other witnesses and his following of leads would be affected. The police investigator withholding belief is preferable in such circumstances.

[33] The case becomes more complicated if we think that the patient's belief in her survival might be influenced by what the doctor believes, say by cues from the doctor's behaviour. Let's assume that the doctor is practised at not behaving in ways that influences patients to believe that they won't survive.

[34] Here is an alternative example: An advanced alien civilisation is discovered on some far away moon and various facts about these aliens are relayed to the general public. Given the significance of the discovery, assuming the testifiers are trustworthy and so on, generally it is appropriate for people not just to believe the facts relayed, but, in normal circumstances, it is appropriate for them to try to gain an understanding of the alien civilization based on the available facts.

[35] That there is an epistemic norm in play in the case of the police investigator might be contested. An alternative claim would be that the police investigator's intellectual agency is governed by a prudential or moral norm. Notice, however, that the reasons provided in footnote 25 as to why it is appropriate that the police investigator withhold belief are precisely epistemic reasons, albeit

The layperson should form her belief based on the eyewitness's testimony, while the police investigator should withhold belief. The police investigator is required to do much more intellectually before intuitively it is permissible for him to believe.

In fact, a virtue theoretic approach provides a fruitful way to explain why norms direct different belief-formations. Aside from being in a position to deny that there's a relevant decision procedure, the virtue theorist can account for those differences by reference to virtues. For the police investigator simply to believe the eyewitness's testimony when the investigation is just under way, isn't what a virtuous agent who is a police investigator undertaking an inquiry would do. What's required of him given his role in investigating the crime, rather, is open-mindedness and intellectual thoroughness. He should be very careful to avoid taking a position, such as belief, that might blind him to modally nearby possibilities. A corollary of this is that he should be intellectually thorough before he does take positions on matters that are subject to his investigation. In this case, this means doing plenty more investigating before believing the eyewitness testimony.[36]

For the layperson, however, matters are different. While it would divert us from the purposes of this paper to focus in a lot of detail on the testimony literature, it's plausible that the layperson should be counterfactually sensitive to certain factors in her reception of the eyewitness testimony. This involves being sensitive to the delivery and content of the testimony such that, had the delivery, say, been provided with suppressed sniggers and the content amounted to a very outlandish claim, then the testimonial recipient wouldn't simply believe the testimony.[37] If this condition is satisfied and the recipient is in a normal epistemic environment, then it is plausibly virtuous for her to trust the testifier. Trusting in such conditions will allow her to avoid missing out on plenty of epistemic goods and to play a part in the circulation

---

with a diachronic aspect.

[36] For a related discussion of how roles can influence the ethics of belief, see Sandford Goldberg, "Should Have Known," *Synthese* 194, 8 (2017): 2863-2894.

[37] For further discussion, see Jennifer Lackey, "Why We Don't Deserve Credit for Everything We Know," *Synthese* 158 (2007): 345–361; Duncan Pritchard, "Knowledge and Understanding," in Duncan Pritchard, Alan Millar, and Adrian Haddock, *The Nature and Value of Knowledge: Three Investigations* (Oxford: Oxford University Press, 2010), 41; Shane Ryan, "A Humean Account of Testimonial Justification," *Logos & Episteme* 5, 2 (2014): 209-219; Shane Ryan, "Virtuous Testimonial Belief in Young Children," *South African Journal of Philosophy* 38, 3 (2019): 263-272; Shane Ryan, Chienkuo Mi, and Masaharu Mizumoto, "Testimony, Credit, and Blame: A Cross-Cultural Study of the Chicago Visitor Case," in *Ethno-Epistemology: New Directions for Global Epistemology*, eds. Masaharu Mizumoto, Jonardon Ganeri, and Cliff Goddard (New York: Routledge, 2020), 94-113.

of epistemic goods within her epistemic environment. In her case then, being attentive and trusting is virtuous.[38]

More complicated cases are possible. While, say, the police investigator's role as an investigator is what is salient in judging the appropriateness of his intellectual response, we typically occupy numerous roles. This means that more complicated cases are possible in which it's less clear that his role as investigator is what is salient. The point here is not to adjudicate between such cases in advance, rather the point is that such roles will have a bearing on how agent's should respond intellectually.[39]

In the second case we again have two agents, and again there is the intuition that how the two agent's should form their beliefs is different. This time though *it's not epistemic norms that seem to be pulling in different directions, rather it's a prudential norm on the one hand and an epistemic norm on the other hand*. The patient has a very good prudential reason to form her belief in such a way as to make it more likely that she will believe that she will survive, even if epistemically it seems she shouldn't.[40] Indeed, the self-concern reflected in a belief-forming process that leads her to significantly improves her chances of survival at the potential cost of false belief seems virtuous. The doctor on the other hand shouldn't form her belief in the same sort of way. Epistemic norms governing experts mean that it wouldn't be virtuous if she believed, contrary to the evidence available, that the patient would survive.[41]

---

[38] John Greco makes a very similar point to the one made with reference to this case, though not through a virtue responsibilist framework (John Greco, "Knowledge, Testimony and Action," in *Knowledge, Virtue, and Action: Putting Epistemic Virtues to Work*, eds. Tim Henning and David P. Schweikard (London: Routledge, 2013), 15-29).

[39] Of course, this is in line with Aristotle's doctrine of the mean, whereby what the mean is will precisely depend on the situation of the of the agent, which allows for the possibility that two different virtuous agents will respond intellectually in two different ways.

[40] One could try to make the case that if epistemic normativity is governed by the goal of maximising true beliefs and minimising false beliefs, then, presuming she is generally a good epistemic agent, even epistemic normativity might demand that she believe she will survive. The reason being that if she survives, then she'll do better vis-a-vis this goal. Again, this is assuming a diachronic dimension to epistemic normativity. For more on such a type of move, see Roderick Firth, "Epistemic Merit, Intrinsic and Instrumental," *Proceedings and Addresses of the American Philosophical Association* 55, 1 (1981): 5-23 and Selim Berker, "Epistemic Teleology and the Separateness of Propositions," *Philosophical Review* 122, 3 (2013): 337-393.

[41] There is a question, as in the other cases, as to exactly which virtues are in play in the case, which I won't go into here to avoid distraction from the main points that the case is intended to illustrate. To do otherwise would require a significant amount of argumentation given that particular accounts of the virtues are absent or only just being developed in the contemporary literature.

The final case shows that *how we should respond intellectually can't be reduced to how we should form our beliefs*. Intuitively, what is required of our intellectual agency goes beyond forming our beliefs in appropriate ways. As the case illustrates, sometimes we are required to seek understanding. In the particular case described, it is plausible that we are required to do so from a moral or all things considered norm, but there may be cases in which we should seek understanding for epistemic reasons that we should seek out understanding. In any case, a primary care-giver who learnt that their child cried inconsolably but did nothing to understand why that would be criticisable for lacking virtue.[42]

Even if one accepts the individual morals taken from the three cases, one may still object to the broader claim that such morals taken together support the view that no decision procedure can be identified. It is helpful to be explicit about what the cases taken together show. If one accepts the response intuitions to these cases, they show that there is no agent-neutral decision procedure. How one should act intellectually precisely depends on specifics relevant to that person, such as their social role. Still, one might wonder whether there might be a decision procedure that takes social roles and other relevant features into account. While this isn't in principle ruled out, it looks an unlikely prospect given the diversity of the features that bear on the ethics of intellectual agency.

Consideration of various epistemic cases undermines the claim that some decision procedure can be found that will tell us what the intellectually right thing to do is in any given situation. There is, however, more to say than this. Just as is the case in virtue ethics with regard to right action, in virtue epistemology with regard to intellectual agency, we can also make the move that although no correct decision procedure is to be found, agents should act intellectually as the virtuous agent would characteristically act in the circumstances. This involves acting intellectually virtuously and not intellectually viciously. In other words, we should act intellectually from intellectual courage, intellectual thoroughness, open-mindedness, inquisitiveness, and so on.

A further dynamic, and one that goes beyond what we find in the analogue account of virtue ethics provided by Hursthouse, however, is that being an intellectually virtuous agent involves not only acting from intellectual virtue generally, but doing so when appropriate. After all, the patient's behaviour in believing that she will survive is intuitively appropriate. The fully virtuous agent is an agent who responds appropriately in cases in which, say, the moral or prudential trumps the epistemic and likewise in cases in which the epistemic trumps the

[42] Understanding is also perhaps required of the police investigator.

moral.[43] I take it that cases of the former are more obvious than cases of the latter. For a case of the latter we can imagine some low level wrong being done, say being late by 15 minutes for a meeting with a friend because one is close to understanding how to solve an important problem in philosophy.[44]

The lesson here is that just as the good moral agent doesn't moralise everything, in that she doesn't act from morality in situations that don't call for a moral response, the good intellectual agent doesn't intellectualise everything, in that she doesn't act from the intellectual in situations that don't call for intellectual response. Again though, it's doubtful that any decision procedure can be drawn up as to when a situation calls for intellectual action and when it doesn't. The result then is that action guidance provided by the virtue theoretic account of intellectual agency offered is limited both as to how exactly one should act intellectually in terms of what virtue should be applied, and as to when one should act intellectually in terms of whether one should act from an intellectual virtue or not.[45]

The guidance offered by the account of the ethics of intellectual agency doesn't stop at the instruction to act as the virtuous agent would act, it extends to instructing us to be open-minded, intellectually courageous, and so on. In other words, we're instructed to act intellectually virtuously. What intellectual virtues such as open-mindedness and intellectual courage involve, however, is something that requires articulation, which is not something that is addressed by this paper. That the role one occupies can bear on what is virtuous in a given situation is also highlighted without being systematically detailed. The account that has been set out here, therefore, isn't intended to be the final word, but rather one of the first words in a virtue theoretical approach to the ethics of intellectual agency—a domain argued to be more appropriate for an ethics in epistemology than the ethics of belief. While the account articulated stakes out a virtue theoretic position, the task of

[43] If there are cases of incommensurability, then it is plausible that virtuous agents may permissibly respond in different ways.

[44] For more on the moral not always trumping the non-moral, see Susan Wolf, "Moral Saints," *Journal of Philosophy* 79, 8 (1982): 419-439; Rinard, "Equal Treatment"; Shane Ryan and Fei Song, "Famine, Action, and the Normative," *Journal of Value Inquiry*, forthcoming.

[45] The reader may have noticed that in both the moral and epistemic domains, the guidance provided by both Hursthouse's virtue ethics and the proposed virtue theoretic ethics of intellectual agency respectively is to act as the virtuous agent would characteristically act. The guidance in the respective domains is not to act as the morally virtuous agent would characteristically act and not to act as the intellectually virtuous agent would characteristically act, as this would be misleading as the above indicates.

providing accounts of the individual intellectual virtues and how virtues and roles interact remains.[46]

4. Conclusion

This paper makes the case that we need an ethics of intellectual agency rather than a mere ethics of belief. It's pointed out that there are aspects of intellectual life that go beyond or are not reducible to doxastic attitudes. What follows is a proposal for just such an ethics of intellectual agency. The claim advanced is that we should act as the virtuous agent would characteristically act in the circumstances. It's acknowledged that such an approach might be criticised for failing to provide action-guidance. The response is that it can provide a degree of action guidance with reference to the virtues but also that it's a mistake to think that any approach can provide a formula that gives accurate guidance in every circumstances.[47]

[46] In fact several recent papers in the literature attempt to do just this. For example, Wayne Riggs and Jason Baehr have provided accounts of open-mindedness as an intellectual virtue, while Nathan King has done the same for perseverance, as have Chienkuo Mi and Shane Ryan for skilful reflection (Wayne Riggs, "Open-Mindedness," *Metaphilosophy* 41, 1 (2010): 172-188; Jason Baehr, "The Structure of Open-Mindedness," *Canadian Journal of Philosophy* 41, 2 (2011): 191-213; Nathan King, "Perseverance as an Intellectual Virtue," *Synthese* 191, 15 (2014): 3501-3523; Chienkuo Mi and Shane Ryan, "Skilful Reflection as a Master Virtue," *Synthese* 197, 6 (2020): 2295-2308). Also, for an account of the intellectual vice of epistemic malevolence, the existence of which maintains a symmetry between virtue theory in ethics and virtue theory in epistemology, see Jason Baehr, "Epistemic Malevolence," *Metaphilosophy* 41, 1 (2010): 189-213.

[47] Thanks to Andrea Robitzsch and several blind reviewers for their feedback on this work.

DISCUSSION NOTES/ DEBATE

TRUE KNOWLEDGE

Peter BAUMANN

ABSTRACT: That knowledge is factive, that is, that knowledge that *p* requires that *p*, has for a long time typically been treated as a truism. Recently, however, some authors have raised doubts about and arguments against this claim. In a recent paper in this journal, Michael Shaffer presents new arguments against the denial of the factivity of knowledge. This article discusses one of Shaffer's objections: the one from "inconsistency and explosion." I discuss two potential replies to Shaffer's problem: dialetheism plus paraconsistency and epistemic pluralism. This is not to be understood so much as a criticism of Shaffer's view but rather as a request to develop his very promising objection from inconsistency and explosion further.

KEYWORDS: knowledge, truth, factivity, dialetheism, paraconsistency, epistemic pluralism, Moore's Paradox, Michael Shaffer

In a concise and very neat recent article,[1] Michael Shaffer presents several new arguments against the denial of the factivity of knowledge, that is, against the denial of the following schema:

(Factivity) S knows that *p* only if *p*.[2]

---

[1] Michael J. Shaffer, "Can Knowledge Really Be Non-Factive?," *Logos & Episteme* XII, 2 (2021): 215-226. For related issues see also: Michael J. Shaffer, "Non-Exact Truths, Pragmatic Encroachment and the Epistemic Norm of Practical Reasoning," *Logos & Episteme* III, 2 (2012): 239-259.

[2] Much of the recent discussion of the denial of (Factivity) within epistemology started with Allan Hazlett, "The Myth of Factive Verbs," *Philosophy and Phenomenological Research* 80 (2010): 497-522. However, there have been precursors, for instance in the philosophy of science: See, e.g., Ilkka Niiniluoto, *Critical Scientific Realism* (Oxford: Oxford University Press, 1999), 84-85. Also before Hazlett, Stjernberg proposed to restrict factivity to certain cases in order to deal with Fitch's paradox, the paradox of the knower and other problems: see Fredrik Stjernberg, "Restricting Factiveness," *Philosophical Studies* 146 (2009): 29-48. In reply to Hazlett see: Daniel Nolan, "Non-Factivity about Knowledge: a Defensive Move," *The Reasoner* 2, 11 (2008): 10-11; John Turri, "Mythology of the Factive," *Logos & Episteme* 2, 1 (2011): 141-150; Savas L. Tsohatzidis, "How to Forget that 'Know' is Factive," *Acta Analytica* 27 (2012): 449-459. See also Hazlett's reply to Turri and to Tsohatzidis in: Allan Hazlett, "Factive Presupposition and the Truth Condition on Knowledge," *Acta Analytica* 27 (2012): 461-478. Michael Hannon and Wesley Buckwalter have invoked the idea of "protagonist projection" against Hazlett and others (in Buckwalter's case based on experimental results): See Michael Hannon, "'Knows' Entails Truth." *Journal of Philosophical Research* 38 (2013): 349-366; Wesley Buckwalter, "Factive Verbs and Protagonist Projection,"

Shaffer discusses those deniers of (Factivity) who claim that one can know some proposition even if it is not strictly true but only approximately true (and thus strictly false).[3] For instance, one can, according to this view, know that the speed of light in the vacuum is 300 000 kilometers per second even if this is strictly speaking false because the speed of light is a bit lower.[4] Shaffer calls this "quasi-factivism about knowledge."[5] The deniers of factivity Shaffer has in mind thus defend the following schema:

(Denial) S knows that *p* only if *p* is true or approximates the truth.[6]

Episteme 11, 4 (2014): 391-409. Later, Buckwalter and Turri offered empirical results favoring a different account, the "representational adequacy account", now critical of factivity: See Wesley Buckwalter and John Turri, "Knowledge, Adequacy, and Approximate Truth," *Consciousness and Cognition* 83 (2020): 102950; see also: Wesley Buckwalter and John Turri, "Knowledge and Truth: a Skeptical Challenge," *Pacific Philosophical Quarterly* 101 (2020): 93-101. For another recent defense of non-factivity of knowledge see Adam Michael Bricker, "Knowing Falsely: the Non-factive Project," *Acta Analytica* (forthcoming).

[3] Is there symmetry such that corresponding to approximate truths which are strictly false we also have approximate falsehoods which are strictly true? Would all strict truths (or only some?) constitute approximate falsehoods? And what should one say about Chuck who is a borderline case of "bald" as well as of "not bald"? Are both statements – "Chuck is bald" and "Chuck is not bald" – neither strictly true nor stricly false but both approximately true and approximately false? I can only raise these questions here but would like to suggest that one should not attribute a 4-valued logic to the quasi-factivist (*strictly true*, *strictly false*, *approximately true*, *approximately false*) or even a fuzzy logic (see Lofti Zadeh, "Fuzzy Sets," *Information and Control* 8 (1965): 338-353). Nothing forces the quasi-factivist to go this way – and they better don't go this way in order to avoid unnecessary complications for their account. See also fn.10 below.

[4] To be sure, the claim here is not that one can know that the speed of light is *roughly* 300 000 kilometers per second; this would be strictly speaking true and constitute "strict" knowledge. There is all the difference between knowing approximately that *p* and knowing that approximately *p*; neither entails the other (though depending on what exactly "approximately" means, they might be compatible with each other).

[5] See Shaffer, "Can Knowledge Really," 218.

[6] We don't have to discuss here what exactly "approximation" means. – If one takes being strictly true as an extreme, ideal case of approximation, one can simplify the above and just claim the following:

(Denial*) S knows that *p* only if *p* approximates the truth.

This seems to be Shaffer's choice. To be sure, Shaffer also remarks that "all approximate truths are false" (Shaffer, "Can Knowledge Really," 220). But this would mean, given that approximate truth is considered to be necessary for knowledge, according to the quasi-factivist (see Shaffer, "Can Knowledge Really," 218-219), that one cannot know strict truths (nor know of a strict falsehood that it is false – an assumption which Shaffer needs later: see Shaffer, "Can Knowledge Really," 220). This, however, would be very implausible. Hence, charity demands that we take Shaffer to

I take Shaffer's most serious objection to this denial to be the one from "inconsistency and explosion."[7] Suppose, as (Denial) allows, that can know some proposition *p* which is strictly false but approximately true. Then a subject S can be in the following predicament:

(1) S knows that *p*,

and

(2) S knows that *p* is false.[8]

Given some principle of closure according to which S can come to know something on the basis of competent inference from something (else) they know,[9] and given the plausible principle (also known to the subject, we may assume) that *p* is false iff not-*p*, we can get from (2) to the following, too:

(3) S knows that not-*p*.

A small step (given (1), (3) and knowledge closure under conjunction introduction) finally leads to the following:

(4) S knows that (*p* and not-*p*).

This is worrisome because we would have to attribute inconsistent beliefs to S, in particular beliefs in contradictions. Assuming (in conformity with classical logic) that contradictions are false, and given (Denial), the quasi-factivist lacks any good reason to criticize S for holding inconsistent beliefs: If one can know falsehoods, then why not contradictions? And if one's belief in some contradiction constitutes knowledge, then what can be wrong with holding it, even if it is inconsistent? Lacking any other reason to oppose inconsistency in general, adherents of (Denial) are thus facing a serious problem. And how could they oppose inconsistency in general if their own position allows for it in so many cases – in all cases of knowledge of approximate truth and strict falsehood?[10]

---

favor a broader notion of approximate truth or knowledge allowing for strict truth or knowledge as an extreme case. For the sake of simplicity of the discussion that follows, I will stick with (Denial) instead of (Denial*) here and use the more narrow notion of approximate truth which excludes strict truth. Nothing substantial depends on this difference here.

[7] Shaffer, "Can Knowledge Really,", sec.3.

[8] "False" means "strictly false" here.

[9] See, e.g., Peter Baumann, "Epistemic Closure," in *The Blackwell Guide to Epistemology*, eds. John Greco and Ernest Sosa (Oxford: Blackwell, 1999), 187-205.

[10] (4) attributes a syntactically inconsistent belief to S. In order to attribute semantic inconsistency, we have to assign truth-values to both "*p*" and "not-*p*." The latter is strictly true, but the former is only approximately true (and strictly false). If we were to accept approximate truth and

There is, according to Shaffer an additional problem for (Denial). Given explosion, that is, the logical principle that anything follows from a contradiction,[11] a subject who believes for some particular proposition *p* that (*p* and not-*p*), and who can acquire new knowledge on the basis of competent deduction from that contradictory belief, could come to know any proposition. Shaffer takes this to say that the subject could know everything and acquire omniscience.[12] However, I think Shaffer should not and need not go that far. I think he only can and only needs to claim that any proposition could be known by the subject (but the subject could not know every proposition out of the infinite set of propositions, given the limits of a human subject's minds and given finite time for making inferences).[13]

All this constitutes a very forceful result and objection against the quasi-factivist. There are at least two potential replies to Shaffer's problem for (Denial). I am presenting these replies not as criticisms of Shaffer's view but rather as a request to develop the objection from inconsistency and explosion a bit further.[14]

First, one could embrace dialetheism[15] and claim that some contradictions are true. This would make it possible to accept inconsistent beliefs and thus respond to the objection from inconsistency. What could be wrong with believing something inconsistent but true? Especially if one could even know it, even in conformity with (Factivity)? I suspect that Shaffer is not in favor of this way out, also because dialetheists typically don't accept just any inconsistency: It is hard to see how they could allow for an inconsistency like "The speed of light in the vacuum is 300 000 kilometers per second and it is not 300 000 kilometers per second."

Closely related to this is the dialetheist response to the threat of explosion: paraconsistency.[16] The basic idea is that at least some contradictions do not allow for explosion and derivation of just anything. However, it would be hard to argue for a

approximate falsehood as truth values (in addition to strict truth and strict falsehood), then we would have to accept truth value gluts in some sense: propositions that are both true and false (in the above case: the latter conjunct perhaps also approximately true). We would thus land directly in quasi-dialetheist territory (see below; see also fn.3 above). However, if we stick with bivalence, we get a straight contradiction and semantic inconsistency.

[11] See Shaffer, "Can Knowledge Really," 221.

[12] See Shaffer, "Can Knowledge Really," 221-222.

[13] So, instead of the stronger claim "For any proposition, the subject knows it" only the weaker claim "For any proposition, it is possible for the subject to know it" would follow (and even "It is possible for the subject to know every proposition" would not follow).

[14] I understand that the brevity of Shaffer's article precluded going into the discussion of such replies.

[15] See, e.g., Graham Priest, "What Is so Bad about Contradictions?," *The Journal of Philosophy* 95 (1998): 410-426.

[16] See, e.g., Priest, "What Is so Bad."

restriction of explosion that could be useful to supporters of (Denial). Even if one can prevent explosion in some cases (like the case of the Liar paradox), one would still have to show that, say, "The speed of light in the vacuum is 300 000 kilometers per second and it is not 300 000 kilometers per second" or any other odd and ordinary contradiction does not allow for explosion either. This would commit one to a very radical and implausible view according to which explosion never or almost never can happen. Alternatively, one could try to find some non-arbitrary way of drawing a principled distinction between pairs of strict falsehoods and approximate truths that allow for some explosion and pairs of strict falsehoods and approximate truths that don't allow for explosion. But then it would be very hard, to say the least, to prevent some ordinary contradictions on the "legal" side of this dividing line from "exploding." The supporters of (Denial) thus face a dilemma here if they try to restrict explosion: A complete or almost complete restriction is implausible and a partial restriction is of no use to them. And one certainly does not want to draw no such distinction because then we would end up with disastrous, unlimited explosion, again. I suspect that Shaffer doesn't want to adopt any such strategy involving dialetheism or paraconsistency but I wonder why exactly.

The second way out of Shaffer's objection from inconsistency and explosion might be more attractive to the defender of (Denial). It is epistemic pluralism about knowledge:

> (Pluralism) There is more than one knowledge relation: for instance, knowledge of strict truths ("knowledge-s") and knowledge of approximate truths and strict falsehoods ("knowledge-a").[17]

According to (Pluralism), we would have to replace (1)-(3) above (relating to some proposition *p* which is strictly false but approximately true) by the following:

> (1*) S knows-a that *p*
>
> (2*) S knows-s that *p* is false
>
> (3*) S knows-s that not-*p*.

There won't, however, be a replacement for

> (4) S knows that (*p* and not-*p*).

Neither

> (4') S knows-a that (*p* and not-*p*)

Nor

[17] – and perhaps some additional knowledge relations: see below.

(4") S knows-s that (*p* and not-*p*)

will do because *knowing-a* does not entail *knowing-s* nor does *knowing-s* entail *knowing-a* (they are incompatible with each other). No conjunction introduction can lead to (4') or to (4'). There will thus be no inconsistency (no contradiction believed) and therefore also no threat of explosion.

Pluralism of this sort (a relative of but distinct from epistemic contextualism[18]) would also offer a solution to the "Moorean objection" Shaffer raises:[19] (1) and (2) entitle S to claim something of the following form:

(5) I know that *p* but *p* is false.[20]

This clearly is infelicitous, and one can imagine an argument to the effect that the infelicity is a semantic one, leading to a contradiction.[21] However, given the pluralism above, all the subject could claim here is something of the form

(5*) I know-a that *p* but *p* is false.[22]

For instance, someone could make the following claim:

(6) I know in an approximative way that the speed of light is 300 000 kilometers per second but that is strictly speaking false.

And something like this doesn't seem infelicitous at all any more.[23]

One can wonder, however, what reason the pluralist can give us still to call what they call "approximate truth" "truth" and still to call what they call

[18] See, e.g.: Stewart Cohen, "How to Be a Fallibilist," *Philosophical Perspectives* 2 (1988): 91-123; Keith DeRose, "Contextualism: An Explanation and Defense," in *The Blackwell Guide to Epistemology*, eds. Greco and Sosa, 187-205.

[19] See Shaffer, "Can Knowledge Really," sec.5. The remaining objection Shaffer offers concerns the safety view of knowledge (see Shaffer, "Can Knowledge Really," sec.4). However, I think this objection has less force because the safety view is very controversial anyway, and for independent reasons.

[20] Alternatively: (5') I know that *p* but not-*p*.

[21] Defenders of (Denial) could object that a semantic explanation using (Factivity) would be question-begging: One would use (Factivity) – which is under discussion here – in order to derive "*p*" from the first conjunct of (5) as well as "*p* is false" from the second conjunct of (5); from the latter one would derive "not-*p*" and then finally put "p" and "not-p" together via conjunction introduction in order to get a contradiction. The contradiction would thus explain the infelicity but only by using the controversial (Factivity). However, defenders of (Factivity) could argue back that the possibility of giving the best explanation of the infelicity of (5) is the semantic one involving (Factivity). This would then give independent support for (Factivity) rather than constitute a case of question-begging.

[22] Keeping in mind that "false" means "strictly false" here.

[23] Again: Knowing in an approximate way that *p* is different from knowing that approximately *p*.

"approximate knowledge" "knowledge." As long as no independent reason is given, this way out is under suspicion of being ad hoc and of constituting a stretch of the semantics of "truth" and "knowledge." Apart from that, what could keep the pluralist from also calling luckily true guesses "knowledge" ("knowledge-l")? What could keep them from starting a huge "inflation" of the set of knowledge relations, perhaps even admitting just anything (any belief?) to the realm of knowledge? It seems much better to replace talk about approximate knowledge by talk about "knowledge-like" states which do not amount to knowledge but are still interestingly close.[24]

I wonder what Shaffer has to say about possible ways out of his inconsistency-and-explosion problem discussed above, especially the pluralist proposal. Answers to the above questions can strengthen his already strong position even further.[25, 26]

[24] – as suggested to me by Michael Shaffer in private communication.

[25] I guess we should also wait for Shaffer's forthcoming book *Quasi-Factive Belief and Knowledge-like States*. – A final side remark. A much more radical form of a denial of factivity says that all falsehoods (including those that are not even approximately true) can be known. Accordingly, knowledge does not require truth. If the "radical denier" also accepts Colin Radford's claim that knowledge does not require belief (see Colin Radford, "Knowledge – By Examples," *Analysis* 27 (1966): 1-11) as well as Crispin Sartwell's claim that knowledge does not require anything beyond truth or belief (see Crispin Sartwell, "Knowledge Is Merely True Belief," *American Philosophical Quarterly* 28 (1991): 157-165), then one ends up with the claim that knowledge does not require anything (one could, of course, still say that knowledge requires knowledge but this would be trivially empty). Thus, one gets a different kind of "explosion:" The word "knowledge" would apply to everything and thus not refer to knowledge anymore.

[26] I would like to thank Michael Shaffer for discussion and comments.

KUHN, VALUES AND ACADEMIC FREEDOM

Howard SANKEY

ABSTRACT: For Kuhn, there are a number of values which provide scientists with a shared basis for theory-choice. These values include accuracy, breadth, consistency, simplicity and fruitfulness. Each of these values may be interpreted in different ways. Moreover, there may be conflict between the values in application to specific theories. In this short paper, Kuhn's idea of scientific values is extended to the value of academic freedom. The value of academic freedom may be interpreted in a number of different ways. Moreover, there are other values which play a role in the functioning of our academic institutions. As with the possible conflict between scientific values, there may be conflict among the academic values.

KEYWORDS: Thomas S. Kuhn, values, theory-choice, academic freedom

1. Introduction

Thomas S. Kuhn was an influential contributor to the history and philosophy of science. He might well be regarded as one of the founders of that interdisciplinary area of study. Kuhn is best-known for *The Structure of Scientific Revolutions*, which introduced the idea of scientific paradigms, and proposed a model of revolutionary scientific change.[1] In his less well-known book, *The Essential Tension*, Kuhn offered insight into the nature of scientific values that may be relevant to current discussion of academic freedom.[2] Kuhn's insights into scientific values apply to the domain of scientific theory-choice. But they have interesting application to the question of academic freedom.

2. The Role of Values in Theory-Choice

Kuhn provoked the ire of philosophers of science who took him to defend an irrationalist conception of the choice between competing scientific theories. He compared the choice between such competing theories (or paradigms) as geocentric and heliocentric astronomy or phlogistic and oxygen chemistry to a shift of visual

[1] Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 4th ed. (Chicago: University of Chicago Press, 2012).

[2] Thomas S. Kuhn, *The Essential Tension* (Chicago: University of Chicago Press, 1977).

gestalt. He likened the intellectual process that a scientist goes through in making such a choice to a religious conversion experience. He argued that scientific observation is so laden by theory that it is as if scientists in competing paradigms live in different worlds. When they look at the same object, they do not see the same thing.

Kuhn was taken aback by the philosophical criticism of the views he had proposed in *Structure*. In a chapter of *The Essential Tension*, Kuhn sought to assuage the philosophical ire by affirming that his views were less iconoclastic than had been thought.[3] There is a shared set of criteria which scientists employ in the appraisal of competing theories. The criteria guide scientists in their choice of theory. These criteria, which Kuhn describes as 'values,' include accuracy, breadth, consistency, simplicity and fruitfulness. Kuhn described the criteria as values because they do not function as rules which dictate or determine the choice of theory. Rather, they are values, which inform and provide guidance, but do not dictate an outcome.

Kuhn made two points about the nature of these scientific values which are of possible relevance to contemporary discussion of academic freedom. First, Kuhn pointed out that even though scientists may share a commitment to a specific value, such as simplicity, they might not understand that shared value in the same way. Simplicity might, for example, be understood as ease of calculation or as involving the postulation of as few entities as possible to explain a phenomenon. Second, Kuhn pointed out that there might be conflict between the values, so that scientists equally committed to the same set of values might diverge with respect to which value they emphasize the most. Simplicity, Kuhn suggested, might favor one theory, while accuracy or breadth of application might favor another.

3. Academic Freedom as Value

Kuhn's suggestions about the values that guide scientific theory-choice may appear to have nothing whatsoever to do with academic freedom. But it is his points about the interpretation of the values and the potential for conflict between them that may have some bearing on the topic of academic freedom. For academic freedom is a cherished value of academic institutions. It might be understood in various ways. And it might conflict with other values that are also deemed important.

There are numerous values that operate within our universities and other academic institutions. The freedom to conduct one's research in a way not constrained by political or other pressure is one such value. The value of truth as an aim of inquiry is another value. Scholarly values, such as the norms that govern the

[3] See Kuhn, *The Essential Tension*, 321-4.

collection, handling and interpretation of evidence provide examples of other values that apply in an academic context. There are also ethical considerations that apply, for example, to experimentation on animals or to the way in which research may be conducted on human beings.

Without attempting to exhaustively catalogue the values that apply in an academic setting we can see already that there is a multitude of such values. How are these to be understood? I have mentioned the value of being able to conduct one's research in a way that is not constrained by political pressure. One might think of that as the value of academic freedom. But equally, one might attempt to articulate the value of academic freedom in some other way, such as the freedom to express one's ideas in a public forum without fear of censure. Or, perhaps, it is a simple matter of being able to choose what one teaches in the classroom.

Here we see Kuhn's point about the multiple understandings of a single value. Academic freedom may be understood in different ways. What about Kuhn's other point about the tension between values? A scientist-colleague once complained to me that his proposal to surgically implant a second heart into an experimental lab rat had been blocked by the ethics committee. The ethical requirement to not unduly cause harm to an animal had been taken to outweigh that scientist's curiosity as to what might happen once the rat had two functioning hearts in its chest. It is tempting to say that this is a case in which the value of unconstrained inquiry has run into a competing value about not causing undue harm to animals. No doubt, there are many further sources of potential conflict between the values that run through our academic institutions.

In extending Kuhn's point to academic values, I am broadening the sense of the term 'value' beyond the sense in which Kuhn employed the term. For Kuhn, the scientific values were criteria of theory-choice which scientists use in their appraisal of theories and as the basis of their choice between theories. In extending Kuhn's point to academic values in general I am speaking about a broad range of items that may be of value in an academic context. Some, such as the norms governing the use of evidence, may be analogous to or even identical with criteria of theory-choice. But, other things, such as freedom of inquiry, the aim of truth and ethical constraints on research are values in another sense. They are not criteria of theory-choice. But I think this extension and broadening is warranted. It is warranted because, like the academic values, the properties of theories picked out by the scientific values are properties on which significant value is placed. Kuhn's scientific values and the academic values are both values in this broader sense.

4. The Upshot

Kuhn offered his remarks about the role of scientific values in theory-choice in response to philosophical critics who accused him of making theory-choice irrational, a matter of religious conversion. He was, I think, attempting to tell us about the nature of scientific rationality. It is not a matter of knockdown argument and conclusive proof. Choice of theory is a deliberative process in which multiple competing considerations must be weighed. Even though scientists may in the end agree in their choice of theory, they need not. And, if they do agree, they need not do so on the basis of the same considerations. Much weighing up and balancing of potentially competing considerations is involved. Rational choice of theory requires deliberative judgement rather than following a clear and simple set of rules.

Kuhn was wont to describe the situation by saying that there is no "neutral algorithm of theory-choice."[4] It is not a matter of following a set of rules which will mechanically determine a single fixed outcome that all scientists will accept. This was one of the reasons that Kuhn described the shared criteria of theory-choice as values rather than as rules. They guide and inform scientific decisions but may not dictate unique outcomes.

What is the relevance of Kuhn's insight into the values that govern theory-choice to the issue of academic freedom? We might recognize that academic freedom is one among a number of values with which academic life is imbued. It is not the only such value. It may be understood in different ways, and it may conflict with other values. But, in the same way that the multiplicity of scientific values tells us something informative about scientific rationality, the multiplicity of academic values tells us something informative about academic life. There is no need for despair. That is how the institution works.

Perhaps one might agree in general but demur on a point of detail. Surely, some values are the most fundamental of all academic values. Perhaps the value of freedom to pursue one's inquiries unhampered by external constraint might be taken as the most fundamental of the academic values. But this would be like taking one of Kuhn's scientific values as the single most important value. You can do that, and you could set up an algorithm, such that a unique outcome of theory-choice is ensured. But, if you do that, you set up one value as sacrosanct, and you beg the question against the other values. It is not clear that much will be gained by attempting to set up one value as the supreme academic value that outweighs all others. And something might be lost. After all, what brings energy and life to

[4] Kuhn, *Structure*, 198.

academic discussion is often the fact that a myriad different thoughts and approaches engage in vibrant interplay as we explore and develop our ideas and those of others.

REVIEWS

Kirk Lougheed, *The Epistemic Benefits of Disagreement*. Series: Studies in Applied Philosophy, Epistemology and Rational Ethics (Cham: Springer, 2020).

Reviewed by Octavio García

The epistemology of disagreement focuses on two views on how agents should respond epistemically to disagreement: conciliationism and non-conciliationism. The first view requires agents to revise their beliefs in the face of disagreement, while the second view allows agents to remain steadfast in their beliefs in the face of disagreement. Lougheed's *The Epistemic Benefits of Disagreement* (2020) offers a context-sensitive non-conciliationist view on the epistemic response that agents should have in the face of disagreement. It aims to avoid the consequences of the less fine-grained views on how agents should respond to disagreement, and it recommends non-conciliationism in complex research disagreement contexts and conciliationism in simple cases of disagreement. Lougheed concludes that agents are reasonable to remain steadfast in a belief in complex research disagreement scenarios if there are future epistemic benefits that result from steadfastness. For instance, if a researcher holds a belief *P*, and that belief *P* has future epistemic benefits, as producing more true beliefs, then the researcher can rationally remain steadfast in that belief in the face of disagreement.

Lougheed's (80) Benefits of Inquiry Argument (BIA):

1. If agent *S* reasonably believes that there are future epistemic benefits to be gained from continuing to believe proposition *P* in the face of epistemic peer disagreement within a research context *R*, then *S* is rational to be a non-conciliationist about *P* in the context of *R*.
2. *S* believes *P* within the context of *R*.
3. There is at least one epistemic peer of *S's* who believes *not-P* within the context of *R*.
4. *S* reasonably believes that there are future epistemic benefits to be gained from continuing to believe *P* in the context *R*.

Therefore,

5. *S* is rational to be a non-conciliationist about *P* within the context of *R*.

BIA's soundness requires Lougheed to deny that simple and complex disagreements have the same epistemic conditions. Then, Lougheed rejects the standard account of epistemic peerhood because of its stringent requirements to hold a looser notion of peerhood that applies to complex disagreement scenarios. To hold that it is reasonable for agents to remain steadfast in their beliefs, Lougheed must embrace epistemic consequentialism, the thesis that future epistemic benefits justify agents' current beliefs. Finally, Lougheed posits the Giving Up Principle (GUP) to avoid dogmatic non-conciliationist stances (64).

First, Lougheed (19) argues that much of the literature on disagreement focuses on simple cases of disagreement and erroneously draws lessons from simple cases of disagreement for complex cases of disagreement. Yet, simple and complex cases of disagreement have different traits that yield different epistemic properties. First, simple disagreements are about individual propositions that have straightforward methods of confirmation. Second, parties to simple disagreements are alethic experts assessed by truth-tracking records that share expertise. Third, propositions in simple disagreements are marginal or central to our web of beliefs. Hence, simple disagreement cases yield clear conciliationist or non-conciliationist verdicts depending on how easy it is to check the belief in dispute and whether the belief has a marginal or central place in our web of beliefs. For instance, if one disagrees on 2+2=4, one is disagreeing about an individual proposition central to our web of beliefs that is easy to check. In this case, one ought to remain steadfast in one's belief. Otherwise, belief revision would undermine our entire web of beliefs. In the same way, disagreement on which horse won a head to head race yields a clear conciliationist verdict, for it is about an individual proposition marginal our web of beliefs that is easy to check.

Meanwhile, complex disagreement cases yield unclear verdicts because disagreement is about several beliefs in the middle of our web of beliefs which are not easy to check. First, beliefs in complex disagreements are in the middle of our web of beliefs since they are not mathematical nor perceptual, so it is not easy to have a unanimous verdict on them because they are not marginal enough to be easy to give up nor central enough to jeopardize our web of beliefs. Second, disputed beliefs in complex disagreements are not liable to clear confirmation because both parties' credences are so close that slight differences between them will not confirm one position or another, considering that there are no agreed criteria on what would constitute definitive evidence. Hence, parties to complex disagreement are non-alethic experts since there are no truth-tracking criteria to assess them. For instance, disagreement about the meteorological forecast is not marginal nor central to our web of beliefs. It is in its middle since it depends on mathematical models and

perceptual evidence. So, it is not clear that agents should revise their beliefs or not in the face of disagreement. Also, there is no clear-cut method to confirm meteorological forecasts as in disagreement on perceptual or mathematical beliefs. Finally, there are no clear criteria on what is decisive evidence since both meteorologists' credences would be so close that truth-tracking assessment is not possible, making them non-alethic experts.

However, Lougheed fixes the epistemic significance of disagreement by appealing to a broad notion of peerhood in complex disagreements. Lougheed (52-53) argues that one can have the conciliationist verdicts in complex disagreement scenarios without relying on the standard notion of peerhood through Ballantyne's Meta-Defeater Argument.[1] Ballantyne argues that evidence of the existence of a defeater for believing *P* relying on a body of evidence *E* provides a defeater for believing *P*, and if one has evidence of the existence of a defeater for believing *P* relying on a body of unpossessed evidence and one has no defeater for that defeater, then one has an undefeated defeater for believing *P*. This argument applies to complex disagreement for two reasons. First, disputed views in complex disagreements are not particularly strong. Second, disputant parties to complex disagreements are not aware of all defeater defeaters because there are several propositions in dispute.

Furthermore, Ballantyne's Doubtful Fairness Argument[2] states that if the existence of unpossessed evidence shows that the sample of evidence one has is unfair, then one should decrease one's confidence in *P*. This argument relies on having reasons to disbelieve that evidence one has is a fair sample, or to have reasons to suspend one's judgment on the fairness of the sample one has. The outcome of this argument is that parties to complex disagreement do not have reasons to believe that their evidence is superior to their opponent's evidence. This outcome is stronger against worldviews since agents do not gather evidence for worldviews in ways that ensure sample fairness. Both arguments make it unnecessary that both parties share the same evidence to produce conciliationist verdicts.

Then, Lougheed posits the Skeptical Epistemic Peerhood (SEP), which is a broader concept of peerhood that applies to real-world disagreements composed of isolated propositions or complex disagreements deeply tied to an agent's worldviews. SEP states that two agents are peers if and only if (a) each agent lacks some dispute independent reason to think that each of their bodies of evidence is superior to that of each other of their opponents and (b) each agent lacks a dispute independent

[1] Nathan Ballantyne, "The Significance of Unpossessed Evidence," *The Philosophical Quarterly* 65, 260 (2015): 315-335.

[2] Ballantyne, "The Significance of Unpossessed Evidence," 326-330.

reason to think they have assessed their body of evidence more accurately than each of their opponents. Ballantyne's Meta-Defeater Argument and Doubtful Fairness Argument support (a), while the lack of positive reasons an agent party to a disagreement has to believe that she is better at assessing evidence than her opponent supports (b). This view on disagreement preserves the conciliationist challenge to disagreements in complex scenarios without relying on the strict notion of peerhood and on the assumption that simple and complex disagreements have the same epistemic traits and hence the same epistemic properties.

Given the differences between complex and simple disagreements and the replacement of a strict notion of peerhood for SEP, Lougheed (96) recognizes that BIA requires a consequentialist conception of justification that states that "if believing a proposition leads to true beliefs, regardless of whether or not that proposition is true, then the belief in the proposition in question is epistemically justified." Talbot's argument to support the consequentialist conception of justification is that "epistemic oughts are normative, epistemic oughts have a source, and the source of epistemic oughts is an end that has true beliefs as a necessary component."[3] Hence, regardless of true beliefs being in the future, they provide epistemic oughts. In this way, Lougheed argues that future epistemic benefits, as true beliefs, can make non-conciliationist reasonable in the face of complex research context disagreements. Here, Lougheed uses a crucial distinction between synchronic epistemic reasons and diachronic epistemic reasons. Synchronic epistemic reasons are reasons currently held by agents to believe *P*, while diachronic epistemic reasons are epistemic reasons derived from future epistemic benefits, as future true beliefs, that justify agents to believe *P*. For instance, current evidence that increases the probability of *P* being true is a synchronic reason to believe *P*. Meanwhile, future epistemic benefits, as the probability of gaining more true beliefs, is a diachronic reason to believe *P*. BIA relies on this distinction to preserve a non-conciliationist verdict in complex research disagreement contexts, for agents can reasonably hold their beliefs in such scenarios if they have diachronic epistemic reasons.

Another requirement of Lougheed's BIA is to claim that epistemic values of avoiding false beliefs and holding true beliefs must be in equilibrium. There are four epistemic values at stake. The first one does not allow trade-offs between true and false beliefs, and it requires that agents always avoid false beliefs. The second one does allow trade-offs between true beliefs and trivial false beliefs, so agents can embrace trivial false beliefs to gain important true beliefs. The third one allows

[3] Brian Talbot, "Truth Promoting Non-Evidential Reasons for Belief," *Philosophical Studies* 168, 3 (2013): 599–618, https://doi.org/10.1007/s11098-013-0139-1.

trade-offs between non-trivial false beliefs and valuable true beliefs. Yet, trade-offs are not allowed between non-trivial false beliefs and inconsequential true beliefs. The fourth one allows trade-offs between non-trivial false beliefs and valuable true beliefs. BIA falls under the third view on epistemic values since it admits trade-offs between trivial false beliefs to gain valuable true beliefs. In this way, future epistemic benefits, in the form of valuable true beliefs, justify steadfastness on current beliefs in the face of complex research disagreement contexts since valuable true beliefs give agents diachronic reasons to maintain their views even if agents face peer disagreement.

Finally, BIA requires a restriction to beliefs to avoid dogmatism and epistemic harms in the face of complex research disagreement. Lougheed (64) advances the Giving Up Principle (GUP) that states that "an agent is irrational to continue to believe P when reasons she is aware of ought to convince her that her belief P is mistaken." What are good reasons to believe that she is mistaken depend on whether the disagreement is complex or simple. Also, the rationality of the researcher facing disagreement depends on the rationality of believing in the probability of the epistemic benefits and not on the actual epistemic benefits.

One can object that there is no reason to think that there are future epistemic benefits that follow from remaining steadfast in one's position in complex research disagreement contexts because of the pessimistic induction argument.[4] Even if Laudan's argument attacks the notion of successfulness in the scientific realism/antirealism debate, one can apply his argument to argue against the thesis that scientific theories are true. It would undermine Lougheed's argument since it requires the existence of future epistemic benefits being available in complex research disagreement contexts so agents can reasonably remain steadfast in their beliefs in disagreement contexts. However, Laundan's argument shows that we have no reason to believe that future epistemic benefits exist since most scientific theories are false.

First, Laudan argues that a realist about science must hold that reference is a necessary condition for the truth of scientific theories. For instance, if there is nothing like gens, then the genetic theory cannot be true. In the next step, Laudan puts forward a list of theories that failed to refer, that is, theories that were false and that were once well confirmed and successful. Laudan's list of theories that failed to refer shows that many theories we believed were truth turned out false. In conclusion, current theories we believe are true will be regarded as false in some future time by enumerative induction. Even if this argument aims to show that the

[4] Larry Laudan, "A Confutation of Convergent Realism," *Philosophy of Science* 48, 1 (1981): 19–49, https://doi.org/10.1086/288975.

reference requirement is not a necessary condition for scientific theories to be confirmed and successful, it also can be interpreted as an argument that shows by enumerative induction that theories that we regard as true nowadays will be false in the future. If this argument is sound, then future epistemic benefits are not guaranteed, and remaining steadfast in one's beliefs in a complex research context disagreement is not reasonable.

Second, one can explain this by arguing that the role future epistemic consequences play in epistemic rationality is asymmetrical. Future epistemic consequences can undermine agents' justification for current beliefs, yet it is not clear that future epistemic consequences can justify agents' current beliefs. First, agents have cognitive and evidential limited resources when it comes to knowing truths. Hence, all truths agents know now will turn out to be falsehoods because of the cognitive and evidential limited resources. So, future epistemic benefits, as true beliefs, will turn out to be false beliefs. And if future true beliefs will turn to be false beliefs, then there is no reason to believe that future epistemic consequences can justify steadfastness in one's beliefs in a complex research disagreement context. Lougheed's argument requires future epistemic benefits to be true in a dogmatic way. Otherwise, they could not provide reasons for non-conciliationism in the face of complex research disagreement contexts.

ERRATUM NOTICE

The citation in footnote 8, page 291, from the article "Epistemic Norms, the False Belief Requirement, and Love," authored by J. Spencer Atkins and published in the previous issue of *Logos & Episteme*, contains an error. The citation reads "Georgi Gardiner, 'Risk and Relevance: How the Relevant Alternatives Framework Models the Epistemology of Risk,' forthcoming in *Synthese*." It should read "Georgi Gardiner, 'Relevance and Risk: How the Relevant Alternatives Framework Models the Epistemology of Risk,' forthcoming in *Synthese*."

NOTES ON THE CONTRIBUTORS

Miguel López Astorga is Full Professor and Researcher at the University of Talca (Chile). His last researches focus on the human inferential activity. He is especially interested in the relations that can be found between syntax, semantics, and pragmatics in that activity. He also addresses the possible cognitive potential of some ancient logics, particularly, Stoic logic. Contact: milopez@utalca.cl.

Peter Baumann is Professor of Philosophy at Swarthmore College. His research focuses on epistemology and also on rational decision making. His main historical interests are in 18th century philosophy (Kant, Reid) and in 20th century analytic philosophy. He also likes to stray into ethics or political philosophy. His most recent monograph was *Epistemic Contextualism. A Defense* (Oxford University Press, 2016). Amongst recent articles are: "Brains in Vats? Don't Bother!" (*Episteme*, 2019), "What Will Be Best for Me? Big Decisions and the Problem of Inter-World Comparisons" (*Dialectica*, 2018), "Trolleys, Transplants and Inequality: An Egalitarian Proposal" (*Erkenntnis*, forthcoming), "Defending the One Percent? Poor Arguments for the Rich?" (*The Harvard Review of Philosophy*, 2014). Contact: pbauman1@swarthmore.edu.

Erhan Demircioglu earned his Ph.D. at the University of Pittsburgh in 2011. He is currently an associate professor of philosophy at Koç University, Istanbul. His main research interests lie in epistemology, philosophy of mind, and philosophy of language. Demircioglu has authored papers in these areas for such journals as *Synthese*, *Philosophical Studies*, *Theoria*, *Philosophia*, *Minds and Machines*, and *Acta Analytica*. Contact: erdemircioglu@ku.edu.tr.

Steven Diggin is a PhD student at the University of British Columbia. His current research is focussed on questions about the nature of knowledge and epistemic justification, especially concerning how these relate to questions about rational activity more generally. Contact: steven.diggin@gmail.com.

Octavio García has a Bachelor's degree in anthropology from the Universidad de Costa Rica (2016), a Bachelor's degree in philosophy from the Universidad de Costa Rica (2017), and a Master's degree in philosophy from the Federal University of Ceará (2020). He is currently studying the epistemic consequences of philosophical disagreement. He collaborates with the research group Philosophy, Metaphysics, and Cognition at the Federal University of Ceará (Brazil). He is interested in the history of philosophy, metaphilosophy, methodology of philosophy, and epistemology of philosophy. Contact: garciaoctavio14@gmail.com.

Tsung-Hsing Ho is Associate Professor at the Department of Philosophy, National Chung Cheng University. He writes on diverse topics, including the ethics and normativity of belief, the value of knowledge, virtue epistemology, ethics, AI ethics, and Chinese philosophy (https://philpeople.org/profiles/tsung-hsing-ho). Contact: tsunghsing.ho@gmail.com.

Shane Ryan obtained his PhD from the University of Edinburgh and is currently Associate Professor of Philosophy at Nazarbayev University. His research is on a variety of topics, including wisdom, epistemic environments, and paternalism, and has been published by leading international philosophy journals and university presses. Two of his current projects are to examine how epistemic environments might be designed to facilitate the development of wisdom and how perceptions of epistemic coverage can influence the consumption of fake news. Contact: shaneryan27@hotmail.com.

Howard Sankey is Professor of Philosophy in the School of Historical and Philosophical Studies at the University of Melbourne (Australia). He teaches in epistemology and philosophy of science. He has published on the problem of incommensurability, epistemic realism and scientific realism. He is the author of *Scientific Realism and the Rationality of Science* (Ashgate, 2008), *Theories of Scientific Method: An Introduction* (Acumen, 2007, with Robert Nola), *Rationality, Relativism and Incommensurability* (Ashgate, 1997) and *The Incommensurability Thesis* (Avebury, 1994). More information may be found at https://philpeople.org/profiles/howard-sankey. Contact: chs@unimelb.edu.au.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes 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 promotes 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.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the *Notes to Contributors*.

Contact: logosandepisteme@yahoo.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

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.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

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 postal and email 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 which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-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 submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

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.

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: 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. Copyright & Publishing Rights

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme are* also freely available on Philosophy Documentation Center web page.