volume VIII♦ **issue** 2
2017

Logos & Episteme

an international journal
of epistemology

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

Contact address:

Institutul de Cercetări
Economice şi Sociale „Gh.Zane"

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

www.logos-and-episteme.proiectsbc.ro

ISSN 2069-053

TABLE OF CONTENTS

RESEARCH ARTICLES
Miloud BELKONIENE, What Are Explanatory Virtues Indicative Of?..... 179
Fabio LAMPERT and John BIRO, What Is Evidence of Evidence Evidence Of?..... 195
Michael J. SHAFFER, A Thoroughly Modern Wager..... 207

DISCUSSION NOTES/ DEBATE
Andy MUELLER, Pragmatic or Pascalian Encroachment? A Problem for Schroeder's Explanation of Pragmatic Encroachment..... 235
John N. WILLIAMS, Still Stuck on the Backward Clock: A Rejoinder to Adams, Barker and Clarke..... 243

REVIEWS
Tobias Endres, Pelegrinno Favizzi, Timo Klattenhoff (editors), *Philosophie der Kultur-und Wissensformen. Ernst Cassirer neu lessen*, reviewed by Ioan Alexandru Tofan..... 273

Notes on the Contributors……. 277
Logos and Episteme. Aims and Scope……... 279
Notes to Contributors……..... 281

RESEARCH ARTICLES

WHAT ARE EXPLANATORY VIRTUES INDICATIVE OF?

Miloud BELKONIENE

ABSTRACT: This paper discusses an assumption on which explanationist accounts of the evidential support relation rely with a focus on McCain's recent account. Explanationist accounts define the relation of evidential support in terms of relations of best explanation that hold between the evidence a subject possesses and the propositions she believes. Such a definition presupposes that the explanatory virtues of what best explains a subject's body of evidence is indicative of its truth. Yet, recent cases offered in the literature against McCain's account show that there is no straightforward way of vindicating this assumption.

KEYWORDS: evidentialism, explanationism, explanatory virtues, evidential probability

Offering a satisfying account of the relation of evidential support is one of the main tasks that have to be carried out by philosophers who endorse an evidentialist conception of epistemic justification. This relation holds between a subject's body of evidence and the propositions that receive a certain degree of confirmation from this evidence, and its existence is generally taken, at least by evidentialists, to be a necessary condition for epistemic justification.[1] In other words, evidentialists generally agree on the fact that for someone to be justified in believing that P, P has to be supported by the evidence one has. Given this general agreement, a central question related to the elucidation of the notion of epistemic justification concerns the conditions under which a subject's evidence supports a given proposition.

This paper focuses on accounts of the evidential support relation that define it in terms of relations of best explanation that hold between a subject's evidence and the propositional content of her beliefs. More specifically, this paper offers a critical discussion of McCain's explanationist account of the evidential support

[1] Evidentialists often distinguish doxastic justification from propositional justification. While the relation of evidential support that holds between a subject's evidence and a proposition P is necessary and sufficient for this subject to be propositionally justified in believing P, it is only a necessary condition for this subject to be doxastically justified in believing P. In addition of the relation of evidential support that holds between a subject's evidence and P, doxastic justification requires that a subject's belief that P be properly based on this subject's evidence.

relation.[2] McCain's account happens to have very satisfying results when applied to some problematic cases that have been offered against other explanationist accounts in the literature. Yet, this account, like any other explanationist account of the evidential support relation, relies on a crucial assumption whose plausibility needs to be assessed. This assumption relates to the relation between an explanation's explanatory virtues and its truth. If the relation of evidential support is to be defined in terms of relations of best explanation, then the explanatory virtues of what best explains a subject's evidence has to be indicative of its truth. However, cases offered in the recent literature show that there is no straightforward way of supporting this assumption.

In the first two sections of the present paper, McCain's account is introduced and the assumption concerning the relation between an explanation's explanatory virtues and the truth of this explanation on which it relies is spelled out. In the third section, the plausibility of this assumption is questioned on the basis of a case offered against McCain's account by Byerly and Martin.[3] The fourth section relies on an account of the evidential relevance of explanatory considerations that has been put forward by McCain and Poston[4] and on Leitgeb's[5] theory of rational belief to consider a more sophisticated way for explanationists to vindicate this assumption. In the last section of this paper, I discuss two cases which show that this way of supporting the assumption on which McCain's account relies is ultimately unsatisfactory.

1. Explanationist Accounts of the Evidential Support Relation

According to Evidentialism (*E*), the justification a subject has for believing that a given proposition is true is determined by the body of evidence she has at a certain time. In its strongest form,[6] this thesis can be formulated as follows:

E: S is justified in believing P at *t* iff S's evidence *e* at *t* supports P.

[2] See Kevin McCain, "Explanationist Evidentialism," *Episteme* 10 (2013): 299–315, Kevin McCain, *Evidentialism and Epistemic Justification* (New York: Routledge, 2014) and Kevin McCain, "Explanationism: Defended on all sides," *Logos & Episteme* 6 (2015): 61–73.

[3] T. Ryan Byerly and Kraig Martin, "Problems for Explanationism on Both Sides," *Erkenntnis* 80 (2014): 773–791.

[4] Kevin McCain and Ted Poston, "Why Explanatoriness Is Evidentially Irrelevant," *Thought* 3 (2014): 145–153.

[5] Hannes Leitgeb, "The Stability Theory of Belief," *Philosophical Review* 123 (2014): 131–171 and Hannes Leitgeb, "I—The Humean Thesis on Belief," *Aristotelian Society Supplementary* 89 (2015): 143–185.

[6] A weaker version of it would only define evidential support as a sufficient yet not necessary condition for propositional justification.

Several philosophers have endorsed Evidentialism and proposed different analyses of the notions involved within it. There is, for instance, no agreement among them concerning the nature of a subject's evidence. Likewise, not all evidentialists think that a subject's justification for believing P is determined by the totality of the evidence she has at *t*; some rather consider that justification is only determined by a properly restricted portion of a subject's evidence. Finally, and maybe most importantly, philosophers diverge in the way they conceive the relation of evidential support, central in *E*.

While many tend to conceive this relation within the framework of the Bayesian theory of confirmation – namely in terms of P's conditional probability on a subject's evidence – alternative explanationist accounts have recently emerged. According to Conee and Feldman,[7] the fundamental epistemic principles are those of best explanation, and the conditions under which a belief is propositionally justified by one's evidence are relative to the explanatory relation that holds between its content and one's evidence. The view they suggest, which is referred to by McCain as Best Explanation Evidentialism (*BEE*), has been synthetized by McCain in the following way:

> **BEE**: S, with evidence *e*, is justified in believing P at *t* iff P is part of the best explanation available to S at *t* for why S has *e*.[8,9]

BEE states that evidential support is a matter of explanatory coherence between a subject's evidence and the content of her beliefs and that explanatory coherence can be defined in terms of what best explains this evidence. The explanatory virtues which make an explanation better than another can remain broadly conceived in the context of the present discussion as being typically the explanatory power, the simplicity, the scope and the unificatory force of a potential explanation available to S for why S has *e* at *t*.

McCain offers an account of the evidential support relation which differs from *BEE* with respect to the role played by relations of logical consequence.[10] He believes that defining evidential support in terms of relations of best explanation alone is too restrictive. Cases proposed by Lehrer[11] and Goldman[12] show that S's

---

[7] Earl Conee and Richard Feldman, "Evidence," in *Epistemology: New essays*, ed. Quentin Smith (Oxford: Oxford University Press 2008), 97–99.

[8] *e* should be understood as the subject's total evidence.

[9] See McCain, "Explanationist Evidentialism," 300.

[10] See McCain, "Explanationist Evidentialism," McCain, *Evidentialism and Epistemic Justification* and McCain, "Explanationism: Defended on all sides."

[11] Keith Lehrer, *Knowledge* (Oxford: Oxford University Press, 1974).

evidence *e* can support the belief that P without P making any contribution to the potential explanation of why S has *e*. In these cases, P is only entailed by other propositions that could contribute to explain why S has *e*.[13] To accommodate such cases, McCain suggests that a proposition P available as a logical consequence of the best explanation for why a subject has evidence *e* can be supported by *e* without making any contribution to the potential explanation of why this subject has *e*. He formulates his account, labelled Explanationist Evidentialism (*EE*), as follows:

> **EE**: S, with evidence *e* at *t* is justified in believing P at *t* iff at *t* S has considered P and either
>
> (i) P is part of the best explanation available to S at *t* for why S has *e;* or
> (ii) P is available to S as a logical consequence of the best explanation available to S at *t* for why S has *e*.[14,15]

While *BEE* and *EE* both incorporate the explanationist idea that evidential support should be defined in terms of relations of best explanation that hold between the propositions believed by a subject and the evidence *e* this subject has at *t*, *EE* extends this support to any proposition entailed by the explanation that best explains why this subject has *e* and, because of this, is able to accommodate a larger range of cases.

2. What Is Assumed by Explanationist Evidentialism

Let me first emphasise why *EE* appears as a *prima facie* plausible account of the evidential support relation. Any satisfying account of the evidential support

---

[12] Alvin I. Goldman, "Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Troubles, Reliabilism's Rescue Package," in *Evidentialism and Its Discontents*, ed. Trent Dougherty (New York: Oxford University Press, 2011), 254–280.

[13] See McCain, "Explanationist Evidentialism," 300–305 for complete discussion.

[14] In response to a problematic case put forward by Byerly and Martin in "Problems for Explanationism on Both Sides," McCain offers a new formulation of *EE* amended with respect to its second condition in "Explanationism: Defended on all sides," 339. In this new version, condition (ii) is spelled out in terms of the explanatory consequences of the best explanation for why S has *e* instead of the logical consequences of the best explanation for why S has *e*. According to McCain, P is an explanatory consequence of the best explanation for why S has *e* if and only if P would be better explained by this explanation than ¬P. As formulating condition (ii) of *EE* in terms of explanatory consequences instead of logical consequences has no incidence on the particular issue I will be raising here for *EE*, for the sake of simplicity, I will stick to the original formulation of *EE* in which condition (ii) is spelled out in terms of logical consequences of the best explanation for why S has *e*.

[15] McCain, "Explanationism: Defended on All Sides," 334.

relation should aim at defining it in terms of a known kind of relation that hold between believed propositions and a subject's evidence and whose existence is indicative of the truth of these propositions. In addition, the kind of relations in terms of which evidential support is defined should be able to account for the deductive and the inductive support that propositions can receive from a subject's evidence. Given this aim, relations of best explanation appear as plausible candidates. Firstly, the ubiquity of abductive reasoning in our everyday lives and scientific practices shows that we often take the explanatory virtues possessed by the hypotheses we come to consider to be related to their truth.[16] Secondly, relations of best explanation between sets of propositions and a subject's evidence allows to account for the deductive and the inductive support that propositions can receive from a subject's evidence.

In addition to its *prima facie* plausibility, *EE* is an attractive account because it is potentially illuminating with respect to the aim of belief-attitudes. Truth is commonly regarded as being belief's regulative aim and therefore as being what epistemic justification tracks. Defining the relation of evidential support in terms of relations of best explanation can give us a deeper insight regarding this aim. If evidential support is to be defined as in *EE*, then epistemic justification can be conceived of as not merely tracking truths, but as tracking informative truths. More precisely, if a subject's evidence supports a proposition just in case this proposition is either part of or entailed by an available representation that best explains this evidence, then epistemic justification can be conceived of as tracking truths that are part of potentially informative representations. By potentially informative representations, I mean representations that can potentially provide some degree of understanding of the phenomena that constitute a subject's body of evidence. As the precise relation between firstly states of understanding, secondly belief's regulative aim and thirdly epistemic justification is of utmost epistemological interest and as *EE* appears to be able to provide some ground for its further investigation, there are independent reasons for regarding *EE* as being an attractive account of the evidential support relation.[17]

---

[16] Note that I am not claiming here that the truth of *EE* depends on the validity of such abductive reasoning. I only take the ubiquity of this form of reasoning to explain, at least partly, the intuitive appeal of *EE*.

[17] Several authors have recently focused on the relation that may exist between the distinctive value of knowledge, which is of course related to truth and epistemic justification, and the epistemic value of states of understanding. See for instance Jonathan L. Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003) and Duncan Pritchard, "Knowledge, Understanding and Epistemic Value," *Royal Institute of Philosophy Supplement* 64 (2009): 19–44 and Duncan Pritchard, "Knowledge and

In the present paper, while acknowledging *EE*'s appeal, my aim is to outline a difficulty that is inherent to this account and, more generally, to any explanationist account of the evidential support relation, in order to emphasise the need for a proper response to it. This difficulty relates to the precise relation that exists between an explanation's explanatory virtues and its truth. *EE* relies on the assumption that an explanation's explanatory virtues are somehow indicative of its truth and are thereby indicative of the truth of the propositions that are part of it or entailed by it. This assumption allows *EE* to define the relation of evidential support in terms of relations of best explanation because an explanation's explanatory virtues are not, at least for most of them, intrinsic properties of an explanation but properties that are possessed by it in relation to a subject's body of evidence. Thus, under the assumption that an explanation's explanatory virtues are indicative of its truth, a subject's body of evidence in relation to which an explanation possesses these virtues can be regarded as indicating the truth of this explanation and thereby the truth of the propositions that are part of it or entailed by it. Yet, for *EE* to constitute a satisfying account of the evidential support relation, the plausibility of this assumption, crucial for *EE*, has to be established.

3. No Straightforward Way to Vindicate This Assumption

What I take to be a serious difficulty for supporting the assumption on which *EE* relies has been highlighted in the context of an exchange that took place recently between McCain, Byerly and Martin.[18] Byerly and Martin offered a case designed to show that P being part of the best explanation available to S at *t* for why S has evidence *e* is, in some cases, not sufficient for *e* to support believing that P. As I will argue, the strength of this case comes from the fact that it shows that, at least in some cases, an explanation's explanatory virtues and its evidential probability, namely its probability conditional on a subject's body of evidence, come apart:

> *Sally Case*: Imagine that Sally is the lead detective on an investigation of a burglary. She typically uses an eight-step investigative procedure for crimes of this sort and this procedure involves gathering and analyzing multiple kinds of

understanding," in *Virtue Epistemology Naturalized: Bridges between Virtue Epistemology and Philosophy of Science*, ed. Abrol Fairweather (Dordrecht: Springer, 2014), 315–328.

[18] Originally, Byerly offered a case against *BEE* that involves beliefs about the future in T. Ryan Byerly, "Explanationism and Justified Beliefs about the Future," *Erkenntnis* 78 (2013): 229–243. Then, McCain offered a response to Byerly's case based on *EE* in Kevin McCain, "Evidentialism, Explanationism, and Beliefs about the Future," *Erkenntnis* 79 (2014): 99–109. Byerly and Martin later challenged McCain's response to Byerly's case and offered a new problematic case for *EE* in "Problems for Explanationism on Both Sides" and McCain offered an answer to Byerly and Martin's concerns and new case in "Explanationism: Defended on All Sides".

> evidence – physical evidences, forensic evidences, testimonial evidences, psychological evidences, circumstantial evidences, and so on. Sally is now mid-way through her investigative procedure, having completed four of the eight steps. She has gathered and analyzed the appropriate evidence for these four steps, but has not yet gathered or analyzed evidence that may or may not arise during the final four steps. The list of suspects with which Sally began has been narrowed, and there is one very promising suspect in particular named Jeremy. In fact, the claim 'Jeremy committed the burglary' (call this the Jeremy hypothesis) is the best explanation available to Sally for all of the evidence she currently has obtained through the first four steps. There are multiple witnesses locating someone who fits Jeremy's description at the scene of the crime at the time at which it was committed. Some drug paraphernalia like that which Jeremy commonly uses to feed his drug habit was found at the scene of the crime. Jeremy seems to display a sense of satisfaction or gladness about the robbery. His bank account reflects a deposit shortly after the incident. Other current suspects, while not ruled out, do not fit the evidence Sally currently has anywhere nearly as well as Jeremy does. The Jeremy hypothesis is the best available explanation for the evidence Sally currently has and it is a very good explanation of that evidence.[19]

Byerly and Martin further specify that it often happened to Sally that, after completing the last steps of her eight-step investigation procedure, a new suspect emerged that better fitted the evidence she had gathered. Given this additional fact, they conclude, rightly it seems, that mid-way through her investigation procedure, Sally is not justified in believing that Jeremy committed the burglary, even though it is part of what best explains the evidence she has. Byerly and Martin also note that the Jeremy hypothesis qualifies as a good explanation in this case, namely as an explanation that is explanatory virtuous, and that the fact that Sally is not justified in believing that Jeremy committed the burglary cannot be accounted for on the basis of the poor quality of the Jeremy hypothesis *qua* explanation.

McCain's answer to Byerly and Martin relies on a possible solution that Byerly and Martin consider and ultimately dismiss. This solution consists in arguing that while the Jeremy hypothesis might be the best explanation available to Sally relative to a portion e of the evidence she currently has, it is not the best explanation relative to the totality of the evidence she has, written e^*, which includes her past experiences of investigations. The reason for which Byerly and Martin dismiss this solution is that they do not consider that there is an alternative explanation available to Sally that could explain better than the Jeremy hypothesis why Sally has e^* at t. Contrary to this, McCain argues that there is in fact an

[19] Byerly and Martin, "Problems for Explanationism on Both Sides," 783.

explanation available to Sally that explains her total evidence e^* better than the Jeremy hypothesis.[20] According to what McCain suggests, this explanation consists of a general hypothesis of the form: 'somebody else than Jeremy committed the burglary,' which does not need to single out a particular suspect. In arguing for this, McCain emphasizes what is, in my view, the crucial aspect of the Sally Case:

> Since "It has not at all been uncommon that at these later stages in the process, an alternative suspect emerges who fits the data even better than previous suspects," presumably from Sally's perspective the odds of there being a rival to the Jeremy hypothesis that is as good, or better, of an explanation than the Jeremy hypothesis is at least .5. In light of this, it is plausible that the best explanation of Sally's data (or at least an explanation that is equally as good as the Jeremy hypothesis) is that some currently unconceived hypothesis is correct.[21]

The fundamental problem in the Sally Case is indeed that, given Sally's total evidence, the probability of there being a rival hypothesis which is the correct one is relatively high and, therefore, the probability of Jeremy being the burglar is relatively low. In other words, the evidential probability of the Jeremy hypothesis does not appear to be high enough for Sally to be justified in believing it to be true.[22] However, I disagree with McCain on the fact that this shows that the Jeremy hypothesis is not the best explanation available to Sally mid-way through her investigation procedure. A hypothesis of the form 'somebody else than Jeremy committed the burglary' is not better, *qua* explanation, than the Jeremy hypothesis, when we consider Sally's total evidence e^*; given e^*, it is only as probable, possibly more probable, as the Jeremy hypothesis. The fact that somebody else committed the burglary can neither explain Sally's past experiences of investigations nor the portion of her evidence that could be explained by the

[20] See McCain "Explanationism: Defended on All Sides," 347–348.

[21] McCain "Explanationism: Defended on All Sides," 347.

[22] The notion of evidential probability invoked here relates to the conditional probability of a proposition on a subject's total evidence. There is an ongoing debate concerning this notion among Bayesians relative to the rational constraints on a proposition's evidential probability. While everybody agrees about the fact that the probability of P conditional on e is the probability a subject S who has e should assign to P, according to some, few rational constraints exist on P's unconditional probability and, therefore, P's probability conditional on e strongly depends on S's actual doxastic perspective, namely on the way P's unconditional probability is determined in S's actual perspective. According to others, P's unconditional probability is constrained in such a way that P's probability conditional on e tends to be independent on S's actual doxastic perspective. However, this debate is somehow orthogonal to the point I am discussing here. I will only assume that the evidential probability of P can be defined given S's total evidence and that this probability is the probability S should assign to P given S's total evidence e.

Jeremy hypothesis. Let us suppose that Sally comes to believe that somebody else than Jeremy committed the burglary; would we consider that she gains any understanding of her past experiences of investigation from this? Would we consider that she gains any understanding of the evidence collected on the crime scene? It does not seem to be the case, the reason being that the hypothesis that somebody else than Jeremy committed the burglary cannot, as such, explain these facts and therefore has little explanatory virtues given Sally's total evidence e^* compared to the Jeremy hypothesis.

The Sally Case shows that an explanation's explanatory virtues and its evidential probability can come apart and that therefore, explanatory virtues cannot be taken to be indicative of the truth of an explanation in the sense of supporting assigning a high probability to it. The Jeremy hypothesis is the most explanatorily virtuous hypothesis available to Sally mid-way through her investigation procedure but, given Sally's total evidence, its probability is relatively low.

4. Relations of Best Explanation and the Stability Condition on Rational Belief

The fact that an explanation's explanatory virtues and its evidential probability can come apart entails that, for *EE* to be able to accommodate cases such as the Sally Case, *EE* needs to be amended with the requirement that the evidential probability of a proposition P be sufficiently high when P is part of or entailed by the best explanation for why S has e at t.[23] The issue with the Sally Case is indeed that, given Sally's total evidence, the probability of the proposition 'Jeremy committed the burglary' is too low for Sally to be justified in believing that Jeremy committed the burglary. Consider the following amended version of *EE* which includes a condition relative to P's evidential probability, written $P_r(P \mid e)$:

> **EE***: S, with evidence e at t is justified in believing P at t iff at t S has considered P, $P_r(P \mid e) > x$ where $.5 \leq x < 1$ and either
>
> (i) P is part of the best explanation available to S at t for why S has e.; or
> (ii) P is available to S as a logical consequence of the best explanation available to S at t for why S has e.

[23] McCain himself, in "Evidentialism, Explanationism, and Beliefs about the Future," 106–107, relative to a possible answer he considers to another problematic case offered by Byerly in "Explanationism and Justified Beliefs about the Future" involving justified beliefs about the future, suggests that P's evidential probability plausibly reflects the extent to which e supports P.

The Sally Case no longer constitutes a problem for *EE**. However, one might ask whether relations of best explanation are still required to define the relation of evidential support in *EE** and, more specifically, what an explanation's explanatory virtues are related to if not to its evidential probability. After all, proponents of the Lockean view of rational belief typically hold that it is sufficient, for S to be justified in believing P, that P's probability conditional on S's evidence be higher that a threshold x such as $.5 \leq x < 1$.[24] Thus, one might argue that conditions (i) and (ii) are no longer needed for *EE** to constitute a proper account of the evidential support relation and that the Sally Case actually shows that an explanation's explanatory virtues are not evidentially relevant in the sense that relations of best explanation are not required for belief justification.

To address this worry, one needs to show that relations of best explanation play a particular evidential role that accounts for a crucial aspect of justified belief which cannot be accounted for by the magnitude of the evidential probability of its content alone. A promising first line of argument consists in taking into account an important weakness of Lockean views of rational belief. Consider a fair lottery involving 1000 tickets. A lottery participant who learns that there is a total of 1000 tickets should assign a high probability to the proposition 'ticket m is losing,' where $1 \leq m \leq 1000$, as the evidential probability of the proposition 'ticket m is losing' is high in this case. Now, if this lottery is known to be fair, this lottery participant should also assign a high probability to the proposition 'one ticket is winning.' But if P's evidential probability being higher than a threshold is sufficient for someone to be justified in believing that P, then any lottery participant is justified in believing that each of the lottery ticket is losing and that one of these tickets is winning. Hence, any lottery participant would be justified in holding logically inconsistent beliefs, which in fact appears to be irrational.

According to Leitgeb, the crucial aspect of rational belief that is left aside in Lockean views of rational belief is the stability of the evidential probability of the belief's content.[25] Once this aspect is taken into account, the paradoxical situation which arises when we consider beliefs in lottery propositions can be avoided. In Leitgeb's view, S is justified in believing that P just in case S is justified, given her evidence, in assigning a stably high probability to P. In other words, S is justified in believing that P just in case $P_r(P \mid e)$ is stably high. The notion of stability

[24] See Richard Foley, "The Epistemology of Belief and the Epistemology of Degrees of Belief," *American Philosophical Quarterly* 29 (1992): 111–121 and Richard Foley, "Belief, Degrees of Belief, and the Lockean Thesis," in *Degrees of Belief*, eds. Franz Huber and Christoph Schmidt-Petri (Dordrecht: Springer, 2009), 37–47 for such a view of rational belief.

[25] Leitgeb, "The Stability Theory of Belief" and Leitgeb, "I—The Humean Thesis on Belief."

invoked by Leitgeb has been put forward by Skyrms[26] and relates to the stability of P's probability under conditionalization on new evidence. P's probability is said to be stably high under conditionalization if and only if, when conditionalized on new evidence, P's probability remains high. For his part, Leitgeb proposes to restrict the class of evidential propositions on which P's probability ought to remain stable to the propositions compatible with P's truth.[27] According to him, S, holding evidence *e*, is justified in believing that P just in case the probability S is justified in assigning to P given *e* is higher than a given threshold and remains higher than this threshold when conditionalized on new evidence compatible with its truth.

Leitgeb's account of rational belief constitutes a solution to the lottery paradox. While the probability that a lottery participant is justified to assign to the proposition 'ticket *m* is losing' is high, it is not stably high when conditionalized on new evidence compatible with its truth. If, for instance, a lottery participant learns that all tickets lost except ticket *m* and one other ticket, namely if she learns that only ticket *m* and one other ticket could be winning tickets, which is compatible with the truth of 'ticket *m* is losing,' the probability she should assign to the proposition 'ticket *m* is losing' is .5. Hence, the probability that a lottery participant is justified to assign to the proposition 'ticket *m* is losing' given her current evidence is not stably high. Leitgeb's account shows that Lockean views of rational belief run into paradoxical situations such as cases of fair lotteries because they fail to take into account a crucial aspect of rational belief, namely the stability of the probability one is justified to assign to its content given one's evidence.

If we accept Leitgeb's diagnosis concerning lottery cases and the stability requirement on rational belief he put forward, a plausible way of defending conditions (i) and (ii) of *EE** consists of showing that relations of best explanation that hold between believed propositions and a subject's body of evidence are necessary and sufficient for this subject to be justified in assigning a probability to these proposition that remains stable under conditionalization on new evidence. This is exactly the line of argument taken by McCain and Poston[28] in their

[26] Brian Skyrms, *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws* (New Haven: Yale University Press, 1980).

[27] See Leitgeb, "I—The Humean Thesis on Belief," 163 for a detailed version of his thesis where the class of propositions on which S's probability assignment ought to be stable is defined in terms of the propositions which are possible from the doxastic perspective of the subject, namely propositions that are not believed to be false.

[28] McCain and Poston, "Why Explanatoriness Is Evidentially Irrelevant."

response to Roche and Sober[29] who raise a concern similar to the one just raised regarding *EE**. Roche and Sober argue that if the explanatory virtues of an explanation are evidentially relevant, then the probability of an explanation H conditional on some observation *e* and on the fact that H best explains *e* – written P_r (H | *e*&E) – should be higher that H's probability conditional on *e* alone – written P_r (H | *e*&E). On the basis of the cases they consider, Roche and Sober argue that this is ultimately not the case and that the fact that H best explains *e* adds nothing, in terms of evidential support, to the extent to which *e* makes H more probable independently of H's explanatory virtues.

In their response to Roche and Sober, McCain and Poston argue that while it is true that P_r (H | *e*&E) is not higher than P_r (H | *e*), the probability that S is justified in assigning to H given *e*&E is more stable than the probability that S is justified in assigning to H given *e* alone. They follow Joyce[30] who considers that the stability of a probability assignment is a property of it which reflects the *weight* of a subject's total evidence and argue that the relations of best explanation that hold between the propositions a subject believes and this subject's total evidence reflect the weight of this evidence which is distinct from the probabilistic support that these propositions receive from this subject's evidence. Hence, they conclude, the fact that H best explains *e* makes a substantial difference in the doxastic attitude that S is justified to adopt toward H.

Given Leitgeb's theory of rational belief and McCain and Poston's answer to Roche and Sober's concern, an argument can be made for the assumption on which McCain's account of the evidential support relation relies and for conditions (i) and (ii) of *EE**. Firstly, an explanation's explanatory virtues that are possessed, at least partly, in relation to a subject's body of evidence are not indicative of the truth of this explanation in the sense of being indicative of its high evidential probability; this explains why an explanation's explanatory virtues and its evidential probability can come apart. Instead, they are indicative of the weight of the body of evidence that probabilistically supports the propositions that are part of or entailed by the explanation that possesses these virtues. Secondly, the relation of evidential support cannot only be defined in terms of the probabilistic support that propositions can receive from a subject's body of evidence as the stability of the probability that this subject is justified to assign to these propositions is an essential aspect of rational belief. Since the stability of the

[29] William Roche and Elliott Sober, "Explanatoriness Is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory," *Analysis* 73 (2013): 659–668.
[30] James M. Joyce, "How Probabilities Reflect Evidence," *Philosophical Perspectives* 19 (2005): 153–178.

probability that a subject is justified to assign to a proposition depends on the weight of the evidence that probabilistically support this proposition, the relation of evidential support has to be defined in terms of relations of best explanation that hold between a subject's body of evidence and the propositions believed by this subject.

5. Relations of Best Explanations Are Neither Necessary Nor Sufficient for Stability

The argument that can be made for the assumption on which relies McCain's account and for conditions (i) and (ii) of *EE** heavily depends on the fact that relations of best explanation are necessary and sufficient for a subject to be justified in assigning a stable probability to a proposition. Yet, as it will be shown in this section, there are good reasons to consider that relations of best explanation are neither necessary nor sufficient for a subject to be justified in assigning a stable probability to a proposition.

Roche and Sober,[31] in a response to McCain and Poston, point out that relations of best explanation might not be necessary for $P_r(P \mid e)$ to be stable under conditionalization. They discuss the following case offered by McCain and Poston, which they initially used to show that explanatory considerations affect the stability of a proposition's evidential probability:

> *Exploding Urn Case*: Sally and Tom have been informed that there are 1,000 x-spheres in an opaque urn. Sally and Tom have the same background evidence except for this difference: Sally knows that blue and red x-spheres must be stored in exactly equal numbers because the atomic structure of x-spheres is such that if there are more (or less) blue x-spheres than red, the atoms of all of the x-spheres will spontaneously decay resulting in an enormous explosion. Sally and Tom observe a random drawing often x-spheres without replacement, five blue and five red. The x-spheres are replaced in the urn.[32]

Given the data both Tom and Sally have, they should assign a probability of .5 to the proposition 'the next x-sphere will be blue.' In addition, Sally, contrary to Tom, has a very good explanation for why she observed a drawing of five red x-spheres and five blue. McCain and Poston argue that given the explanation Sally has, the probability she is justified to assign to the proposition 'the next x-sphere will be blue' will remain stable if, say, she observes ten successive drawing of blue x-spheres, while the probability Tom would be justified to assign to this proposition is considerably higher.

---

[31] William Roche and Elliott Sober, "Explanatoriness and Evidence: A Reply to McCain and Poston," *Thought* 3 (2014): 193–199.

[32] McCain and Poston, "Why Explanatoriness Is Evidentially Irrelevant," 149.

As Roche and Sober note,[33] regarding this case, we need to pay attention to what constitutes the explanans and what constitutes the explanandum. The explanation H Sally possesses is relative to what she knows concerning the atomic structure of the x-spheres. In addition, what is explained by H is the random drawing of ten x-spheres she observed. However, the proposition 'the next x-sphere will be blue' to which Sally is justified in assigning a stable probability is neither part of nor entailed by the explanation H that best explains her evidence. What explains the drawing Sally observed surely does not include the proposition 'the next x-sphere will be blue' and does not entail it neither as the atomic structure of the x-spheres does not entail that the next x-sphere will be blue; the next draw will be random. But if the proposition to which Sally is justified in assigning a stable probability is neither part of or entailed by what best explains her evidence, then it is not necessary for P to be either part of or entailed by what best explains a subject's evidence for this subject to be justified in assigning a stable probability to P.

Let me now consider a second case which shows that P being part of or being entailed by what best explains the evidence that a subject has at *t* might not even be sufficient for this subject being justified in assigning a stable probability to P. Consider the following modified version of the Sally Case:

> *Sally Case**: Sally investigates a burglary based on the same procedure as in the original case. During the burglary, a safe has been opened by someone who knew the safe's code. A very promising suspect is Sam who is an employee of the company where the burglary took place and who potentially had access to the safe's code. As in the original case, the Sam hypothesis can explain other pieces of evidence that Sally gathered during her investigation and hence is the best explanation available to Sally as to why she has the evidence she does mid-way through her investigation procedure. However, unlike the original case, it is Sally's first ever investigation and therefore, given her total evidence, the probability of the Sam hypothesis is quite high as Sally has no reason to suspect that a better explanation for her evidence is yet unavailable to her.

Let us assume that, in the Sally Case*, the probability Sally is justified to assign to the proposition 'Sam committed the burglary,' given her evidence, is .8. The question is now to determine if this probability remains high under conditionalization on evidence compatible with its truth, which would be the case if P being part of or being entailed by what best explains the evidence S has at *t* was sufficient for P's evidential probability to be stable. Let us suppose that, mid-way through her investigation procedure, Sally learns that the company's

[33] See William Roche and Elliott Sober, "Explanatoriness and Evidence: A Reply to McCain and Poston," 196–197.

manager's computer was hacked and that 100 other people potentially had access to the safe's code. At this time, Sally knows nothing about these people; they are simply new suspects who have not yet been ruled out. In addition, what she learns is compatible with the truth of the Sam hypothesis as even though these 100 people had access to the safe's code, Sam still could be the burglar. In fact, it is plausible that the Sam hypothesis is still the best explanation available to Sally once she learns this new information given that she knows nothing about the 100 new suspects. Yet, the probability that Sally is justified to assign to the proposition 'Sam committed the burglary' once she has learnt the new hacking information is considerably lower that it was before. This shows that despite the fact that the Sam hypothesis is the best explanation available to Sally, its evidential probability is not stable under conditionalization on propositions compatible with its truth.

Conclusion

When cases such as the Exploding Urn Case and the Sally Case* are considered, the claim according to which relations of best explanations are necessary and sufficient for a proposition's evidential probability to be stable appears to be doubtful. Consequently, the possible argument for the assumption on which McCain's account relies that has been put forward in the fourth section of this paper does not appear as a viable strategy for explanationists.

To overcome the challenge arising from the fact that an explanation's explanatory virtues and its evidential probability can come apart, explanationists should therefore either look for a property of explanations related to their truth that is always possessed by explanatory virtuous explanations, or they should identify another aspect of rational belief that can be accounted for only in terms of relations of best explanation. Both of these options should be thoroughly investigated as *EE* has many theoretical advantages to offer once the crucial assumption on which it relies is properly vindicated.[34]

[34] ACKNOWLEDGEMENT: Many thanks to Gianfranco Soldati, Fabian Dorsch (1974-2017), Marcel Weber, to the members of the Fribourg EXRE research group, to the members of the Geneva Biological Interest research group and to the audiences of the Paris European Epistemology Network meeting for helpful discussions and comments on previous versions of this paper. This publication was made possible through the support of a grant from the Swiss National Science Foundation.

WHAT IS EVIDENCE OF EVIDENCE EVIDENCE OF?

Fabio LAMPERT and John BIRO

ABSTRACT: Richard Feldman's well-known principle about disagreement and evidence – usually encapsulated in the slogan, 'evidence of evidence is evidence,' (EEE) – invites the question, what should a rational believer do when faced by such evidence, especially when the disagreement is with an epistemic peer? The question has been the subject of much controversy. However, it has been recently suggested both that the principle is subject to counterexamples and that it is trivial. If either is the case, the question of what to do in the face of evidence of evidence becomes less pressing. We contend that even if one or the other of these suggestions is right about (EEE) as a general principle about evidence, they leave it untouched insofar as it plays a role in the debates about the rational way to respond to disagreement and, in particular, to disagreement by an epistemic peer. This is because in such cases the evidence about which one has evidence and which is supposed to provide evidence against one's belief is the mere fact of someone's disagreeing, rather than something that is related to the content of the proposition about which the parties disagree. We go on to argue that, so understood, the principle is false.

KEYWORDS: peer disagreement, evidence, Richard Feldman

Richard Feldman's[1] well-known principle about evidence – usually encapsulated in the slogan, 'evidence of evidence is evidence,' (EEE) – invites the question, what should a rational believer do when faced by such evidence? The question has been the subject of much controversy. However, Branden Fitelson[2] has recently claimed that the principle is subject to counterexamples, and even more recently, Comesaña and Tal[3] have argued that it is trivial, and for that reason not subject to counterexamples at all. If either party is right, the question of what to do in the

[1] Richard Feldman, "Reasonable Religious Disagreements," in *Philosophers Without Gods: Meditations on Atheism and the Secular*, ed. Louise Anthony (Oxford: Oxford University Press, 2007), 194-214, "Evidentialism, Higher-Order Evidence, and Disagreement," *Episteme* 6, 3 (2009): 294-312, "Evidence of Evidence is Evidence," in *The Ethics of Belief*, eds. Jonathan Matheson and Rico Vitz (Oxford: Oxford University Press, 2014), 284-300.

[2] Branden Fitelson, "Evidence of Evidence is not (Necessarily) Evidence," *Analysis* 72, 1 (2012): 85-88.

[3] Juan Comesaña and Eyal Tal, "Evidence of Evidence is Evidence (Trivially)," *Analysis* 75, 4 (2015): 557-559.

face of evidence of evidence becomes less pressing. We contend that the arguments in both Fitelson and Comesaña and Tal, even if they apply to (EEE) generally, leave it untouched insofar as it plays a role, as it often does and was intended by Feldman to do, in the debates about the rational way to respond to disagreement and, in particular, to disagreement by an epistemic peer. This is because the evidence that is supposed to provide evidence against one's belief is the mere fact of someone's disagreeing, rather than something that is related to the content of the proposition about which the parties disagree. Here we aim to do two things. First, to show that (EEE), when properly understood as restricted to cases of disagreement, escapes the recent criticisms of both Fitelson and Comesaña and Tal. Second, to argue that this restricted version of the principle is false when applied to disagreement, both peer or non-peer, for reasons quite different from those suggested by either Fitelson or Comesaña and Tal...

1. Fitelson and Comesaña and Tal on (EEE)

Fitelson's rendition of Feldman's principle is as follows:

> (EEE3) If S_1 possesses evidence (E_1) which supports the claim that S_2 possesses evidence (E_2) which supports p, then S_1 possesses evidence (E_3) which supports p.[4]

He then asks us to imagine that a card c is randomly picked from a standard deck and shown to John. Jim knows only the following about c (apart from knowing that John knows which card c is, and the logical consequences of that and (E_1)):

> (E_1) c is a black card.

And let (E_2) and p be the following:

> (E_2) c is the ace of spades.
>
> (p) c is an ace.

Fitelson argues that while Jim has evidence (E_1) which supports the claim that John has evidence (E_2) which supports p, Jim does not have evidence (E_3) supporting p. A plausible principle concerning evidential support says that E (evidentially) supports p if and only if E raises the (epistemic) probability of p.[5]

[4] Fitelson earlier rejects two other formulations of the same principle for similar reasons. (EEE3) is the version found in Richard Feldman, "Evidence of Evidence is Evidence," *Keynote Lecture at Feldmania: A Conference in Honor of Richard Feldman*, UT San Antonio, February 19th (2011).

[5] Cf. Fitelson, "Evidence of Evidence is not (Necessarily) Evidence," 86.

Since Jim's evidence (E_1) is (epistemically) probabilistically irrelevant to p, (EEE3) is false. Fitelson then asks what evidence (E_3) could there be such that (i) Jim has it, and (ii) it supports p and finds none.

Comesaña and Tal argue that there is a simple way to respond to Fitelson. Suppose that (E_3) is the following:

> (E_3) c is not the Jack of hearts.

Since (E_3) is entailed by (E_1), and Jim knows the logical consequences of both the fact that John knows which card c is and the fact that c is a black card, Jim knows (E_3). Since (E_3) raises the (epistemic) probability of p, (E_3) supports p. Therefore, there is evidence (E_3) such that (i) Jim has it, and (ii) it supports p.

Moreover, according to Comesaña and Tal, Fitelson's putative counterexample should fail anyway. For if one has evidence that q, one thereby has evidence of q *or* p, *which* supports p. Hence, there cannot be any counterexample to (EEE3), for it is trivial.

Note that neither Fitelson nor Comesaña and Tal are concerned with the role (EEE) may have in responding to disagreement. *Their* disagreement is over whether (EEE) can be used to form and justify a belief about p. Jim and John are not described as disagreeing. There is, however, a special use of (EEE), as is suggested by Feldman himself, as a principle to be employed in cases of disagreements generally and peer disagreement specifically. While (EEE) has undergone different formulations at Feldman's hands, what these different formulations have in common is that, when applied to cases of peer disagreement, what one of the supposed peers has evidence of is *solely* the fact that the other has evidence. He does not have evidence of *what* evidence the other has. This is a non-trivial restriction, one that puts constraints on the evidential content governed by the principle and thereby on its relation to what the disagreement is about. We will show, first, that so understood, the principle can be shown to be immune to the criticisms both by Fitelson and by Comesaña and Tal. We will then argue that it is false.

In Fitelson's case, Jim has evidence of what evidence John has, that is:

> (F) Jim has evidence (E_1) (c is a black card) which supports the claim that John has evidence (E_2) (c is the ace of spades) which supports p (c is an ace).

(F) says that Jim has evidence that John has evidence (E_2), namely, that c is the ace of spades, which supports p. This is to say more than just that John has *some* evidence (E_2) which supports p, and certainly more than that the evidence Jim has is only that John has *some* evidence. Yet the latter is all (EEE) is supposed to provide to one confronted by disagreement, especially by disagreement by a peer.

(EEE) can be read in two ways, one where the evidence of evidence is of what the latter is, the other where the evidence is merely that there is some evidence. Both Fitelson and Comesaña and Tal read (EEE) in the first way. As we will see, those who appeal to (EEE) in discussions of disagreement, especially peer disagreement, are explicit that it is the second reading that is relevant in that context.

In Fitelson's supposed counterexample, John has evidence that Jim does not have. What that evidence is (= *c* is the ace of spades) is crucial to Fitelson's case against (EEE). The response by Comesaña and Tal also turns on what the evidence of which there is evidence is. But, as we have noted, in cases of disagreement, (EEE) is supposed to be relevant regardless of what the evidence of which there is evidence is. This is so in cases of near-peer disagreement, where the parties have, and see themselves as having, similar but not identical evidence and at least roughly equal epistemic abilities.[6] And it is especially important in cases of peer disagreement, where it is assumed that the two subjects have the *same* body of evidence. This is an important part of what makes them peers (the others being having identical cognitive abilities in general and expertise with respect to the subject in dispute in particular).

A similar argument applies to Comesaña and Tal's response to Fitelson. Assuming that (E_3) is "*c* is not the Jack of hearts," which is logically entailed by (E_1), and the fact that Jim knows the logical consequences of both the fact that John knows which card *c* is and the fact that *c* is a black card, we have not only that Jim knows (E_3) – thereby refuting Fitelson's counterexample – but also that Jim has evidence John does not have. Again, all this turns on what (E_2) is, and, again, John and Jim are not peers.[7]

What about the claim that (EEE3) is trivial? We agree that taking the slogan *evidence of evidence is evidence* without qualification does render it trivial, as Comesaña and Tal argue. But, again, the principle as applied to cases of disagreement restricts the evidence one has to the fact that the other has evidence. On such a reading (EEE3) says that if S_1 has evidence (E_{1^*}) that his disputant S_2 has evidence (E_{2^*}) against *p*, then S_1 has evidence (E_{3^*}) against *p*. It remains to be shown that the principle so understood is trivial.

[6] If the last condition is not satisfied, there will be no pressure at all on them to take the others' position into account, rather than adopting or dismissing it according to whether they see themselves as epistemic inferiors or epistemic superiors.

[7] That they are not peers is clear since one party knows which card *c* is and the other does not. It should also be noted that my having evidence that you have (what you take to be) evidence *eo ipso* renders us non-peers, strictly speaking: even if you do have the evidence I have evidence you have, *you* do not have the evidence I have that you have it.

2. The Correct Reading of (EEE)

Two of Feldman's formulations of (EEE) are:

> (1) If *S* has evidence for the proposition that evidence exists in support of *p*, then *S* has evidence for *p*.[8]

and

> (2) If *S* has evidence (E_1) supporting the proposition that there is someone who has evidence that supports *p*, then *S* has some evidence (E_2) that supports *p*.[9]

Although (1) and (2) may appear to be equivalent, their antecedents are different and that of (1) does not entail that of (2). There being evidence does not entail that someone has it. Suppose the butler did it in the billiard room and subsequently tossed the dagger into the lake. Even though there is evidence that the butler is guilty (the murder weapon with his fingerprints all over it), it is possible that no-one will ever find it. There is evidence, moreover, that Feldman thinks that (2) is a better way to capture what he has in mind. In an attempt to explain (1), Feldman says

> (2*) This does not mean that if I learn that you have evidence for P, I thereby obtain your evidence. If there is experiential evidence, then when you have a headache, you have experiential evidence supporting the proposition that you have a headache. When you tell me that you have a headache, I don't thereby get your headache. But I do then have reason to think that you have a headache. I get evidence, but not your evidence.[10]

The idea is that "when one learns that another person has evidence supporting a proposition, one has evidence supporting that proposition oneself."[11] It is my having evidence that you have evidence that *p* that gives me evidence in favour of *p*, not having the evidence you have. The evidence I get is *solely* about your attitude towards *p*. This is what lies behind what Feldman calls "the key evidential fact about disagreement:"

> (3) If *S* believes *p*, and *S* learns at *t* that an epistemic peer with respect to *p* who shares *S*'s evidence concerning *p* disbelieves *p*, then *S* acquires some evidence against *p* (…) *It is simply learning of a peer who disbelieves p that provides evidence for him against p.* And the

[8] Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 308.
[9] Feldman, "Evidence of Evidence is Evidence," 292.
[10] Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 309.
[11] Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 308.

> underlying idea here can be put not as a principle about justification at all but instead as a principle about epistemic support … (our emphasis)[12]

This fact underlies the following principle about evidential support in cases of peer disagreement:

> (4) The proposition that *S*'s peer - whose evidence concerning *p* is the same as *S*'s - disbelieves *p* is evidence against *p*.[13]

Note that while (1) and (2) are not restricted to *peer* disagreement, (3) and (4) are. It is the fact that *S*'s *peer* disbelieves *p* that occupies the place of 'evidence$_2$' in the principle 'evidence$_1$ of evidence$_2$ is evidence$_3$.' The principle does not say that *S* thereby has access to a proposition on which the peer's disbelief is based. In a case of peer disagreement, *S* and her peer share one and the same body of evidence, *E*. When *S* learns that her epistemic peer disagrees with her, *S* acquires evidence$_1$ supporting the claim that her peer disagrees with her. But it cannot be that *S* learns that her peer disagrees with her by acquiring new information about her peer's evidence, or by learning that her peer has access to a piece of evidence that *S* does not. If that were the case, they would not be peers in the first place. Their evidence would not be the same to start with, thus we would not be facing a problem of peer disagreement. The only new information, or evidence, that *S* has is *that her peer disagrees with her*. Evidence$_1$ is solely about the propositional attitude of *S*'s peer towards *p*. Feldman's formulations in (3) and (4) clearly say that knowing that one's peer denies *p* is sufficient to give one evidence against *p*.[14]

Feldman is not alone in endorsing (EEE) for cases of peer disagreement. When stating the main motivations for conciliationism, David Christensen claims that "the peer's disagreement gives one evidence that one has made a mistake in interpreting the original evidence, and that such evidence should diminish one's confidence in P."[15] Observe that, again, (EEE) is applied to a special case: it is not

[12] Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 298.

[13] Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 298.

[14] Of course, once I come to think, rightly or wrongly, that my supposed peer disagrees with me, I can no longer think her my peer. Thinking that we disagree entails thinking that either we have different evidence or we are evaluating the same evidence differently because of a difference in our epistemic abilities. And not only can I not think that she is my peer – she cannot be, since even if I am mistaken in thinking that she disagrees with me, she does not have the evidence on the basis of which I formed my false belief that she disagrees. For more on these aspects of the matter, see John Biro and Fabio Lampert, "'Peer Disagreement' and Evidence of Evidence," unpublished manuscript.

[15] David Christensen, "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass* 4, 5 (2009): 757.

evidence of *what* evidence your peer has concerning *p*, but evidence solely *that* your peer disagrees with you that gives you evidence against *p*.[16] Thomas Kelly says that our total evidence in a case of peer disagreement includes "(i) the original, first-order evidence *E*, (ii) the fact that you believe that *p* on the basis of *E*, and (iii) the fact that I believe that not-*p* on the basis of *E*."[17] In setting up what she takes to be the problem of peer disagreement, Catherine Elgin claims that "If someone with the same evidence, training, background knowledge and reasoning abilities came to the opposite conclusion from Jack's, *that* is evidence that Jack's grounds are inadequate."[18] (our emphasis) Adam Elga asks, "How much should this news [that your peer disagrees with you] move you in the direction of her view?"[19] Finally, Harvey Siegel claims that at least according to Feldman, Christensen, and Kelly, "the fact that a peer disagrees with one constitutes evidence that is relevant to the rationality of one's belief."[20] They all endorse Feldman's principle as applied to peer disagreement. None of them, however, need to endorse (EEE) in its general form, where the evidence I have evidence of your having is not restricted to the mere fact that you disagree with me.

We have seen that both Fitelson and Comesaña and Tal interpret (EEE) as saying that evidence$_1$ is about what evidence$_2$ is and, in particular, about the latter's relation to *p.* As we have argued, insofar as (EEE) is supposed to be relevant to disagreement, it is not in virtue of these. This is especially clear insofar as it is supposed to be relevant to peer disagreement. If the evidence which I have evidence you have (evidence$_2$) is evidence I do not have (as is the case with Fitelson's Jim), that is enough to render us non-peers, so that in a case of peer disagreement what evidence$_1$ is evidence of cannot be *that.* It must be admitted that the slogan *evidence of evidence is evidence* may encourage thinking that it is (as may (1) and (2)). It would have been less misleading, if less catchy, to say that evidence of disagreement is evidence, thereby distinguishing it from a principle governing evidence in general.[21] In Feldman's different formulations it is clear

[16] The title of Christensen's – "Disagreement as Evidence" – is revealing.

[17] Thomas Kelly, "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology*, *Vol.* 1, eds. Tamar S. Gendler and John Hawthorne (Oxford: Oxford University Press, 2005), 190.

[18] Catherine Elgin, "Persistent Disagreement," in *Disagreement*, eds. Richard Feldman and Ted A. Warfield (Oxford: Oxford University Press, 2010), 54.

[19] Adam Elga, "Reflection and Disagreement," *Noûs* 41, 3 (2007): 484.

[20] Harvey Siegel, "Argumentation and the Epistemology of Disagreement," *Cogency* 5, 1 (2013): 144-145.

[21] It should be stressed that this would by no means make Feldman's thesis circular. As we have shown above, most writers *do* take peer disagreement as evidence against one's belief. What

that what (EEE) says in cases of peer disagreement is that evidence1 is evidence *that* one's supposed peer has a propositional attitude, not evidence about what proposition the attitude is an attitude to. That is the only principle relevant to cases of (supposed) *peers* disagreeing. Not only that, (EEE) understood *a la* (3) and (4) is also the only principle that could be plausibly thought of as a guide to epistemic conduct in any case of disagreement in which one does not know why the other disagrees. If one does know that, normal standards of evidence assessment kick in.

3. Is Evidence of Evidence Evidence?

What about (EEE) as a principle applied to disagreement in general? Is my having evidence that you have evidence that *p* evidence for me that *p*? We do not think that so. Suppose that I suspect you of being the embezzler. Seeing you stealthily burn some papers, I have evidence that you have what you think (or, perhaps, want me to think you think) is evidence that you are the embezzler and are trying to get rid of it. This may give me good reason to suspect that you are guilty. But it is not evidence that you are. Only if I saw the papers and saw that they did, indeed, incriminate you, would I have *that*. But then it would not be the evidence I have that you have evidence that is my evidence that you are. It would be the evidence you have. While in seeing you burn the papers I acquire evidence that you think that the papers are incriminating, that is not evidence that they are. Only if I saw you stealthily burn some papers that I had reason to believe actually contained information incriminating you, would I have evidence that you are guilty. Since I do not know what is in the papers, I do not have such evidence.

Evidence is usually taken to be an indicator, mark, or sign, that something is the case.[22] My frowning is a sign of my anger, the doorbell's ringing an indication

---

they worry about is what, given this, one should do when one learns that one's peer disagrees. Here is where the literature branches into conciliationism, equal weight, stick to your guns, and so forth, these being different views on what is the proper epistemic response to what one should do in the presence of disagreement by a (supposed) peer, given one's own evidence for believing what one does.

[22] In a fallibilist sense. Otherwise, someone's having evidence that a state of affairs obtains would guarantee its obtaining. It is worth noting that, in a different paper, Tal and Comesaña think of evidence in this way: "We will assume … that someone has a proposition as evidence only if that proposition is true …. To make that factivity transparent, we will symbolize that subject *S* has evidence *e* with 'T(*e*) ^ S(*e*).'" (Eyal Tal and Juan Comesaña, "Is Evidence of Evidence Evidence?" *Noûs* 50, 4 (2015): 98) This is why they can say that they "… assume that the proposition that there is evidence for *p* is itself evidence for *p*." (Eyal Tal and Juan Comesaña, "Is Evidence of Evidence Evidence?" 110, footnote 7) (What they mean, presumably,

that someone is at the door, hoof-prints, that the animal went by. We must ask, then, what does my seeing you stealthily burn some papers indicate? Is it a sign of your guilt? A sign that you are the embezzler? Or is it indicating only that *you think* that the papers contain incriminating information? Writers on peer disagreement typically do not say much about what notion of evidence is in play, taking it for granted that learning that one's peer disagrees with one is evidence (in some sense) against one's belief. Perhaps this is because they think that having such evidence comes to nothing more than having a reason not to believe what one does.[23] But why think that someone's believing that *not-p* is a reason (let alone a *good* reason) to believe that *not-p*, unless one thinks that the someone in question is epistemically superior? Even then, having a reason to believe that *p* is not the same thing as having information that indicates that *p*. Take, again, the case where I see you stealthily burn some papers. Does my seeing you burn the papers give me reason to believe you are guilty? Yes, especially if I already suspect you. But it does not indicate that you are. It is no evidence of this.

Insofar as I am not aware of the content of those papers, I cannot say whether they incriminate you. If they do, then you are in fact burning evidence that you are guilty. If they do not, then you are not. There is nothing that tells me which is the case. So, what is my seeing you burn some papers evidence of? Obviously, only that you *take* them to be evidence of your guilt.[24] Here, it is unclear whether Fitelson's notion of evidence as raising epistemic probability would be of any use. Since I do not know what is in the papers, if my seeing you stealthily burn some papers raises the epistemic probability of anything at all, it would be of the fact that you *take* the papers to be incriminating, not of their actually being so.[25] For the latter, I must know what they say. We can agree that evidence is a sign; the question is, of *what* it is a sign?

Nevertheless, it may still be urged that my seeing you stealthily burn some papers *is* evidence, when added to my suspicion of your being the embezzler, that you think they show you to be guilty, which raises the epistemic probability that you *are*. Factoring in my suspicion in this way, will not, however, help the friend of (EEE). What is supposed to raise the epistemic probability that you are guilty is information about you. However, my being suspicious is not information about

---

is that if there is evidence in favor of *p* that entails that *p*, to have evidence that there is is to have evidence that *p*.)

[23] See, for instance, Feldman's quote in (2*).

[24] It need not be good evidence, of course. You may be laying a false trail in order to save the real culprit.

[25] And even less of your being guilty: there can be incriminating evidence against someone innocent.

you, but a fact about me. Suppose that after the evidence is presented to the jury we say "moreover, we have always suspected the defendant." Surely, we do not thereby add to the evidence against him.

Finally, it remains to be explained what it is for something to be a sign, or evidence, that something is the case. We have argued that seeing you burn the papers is a sign only that you take the papers to be incriminating, not that they are. Is there a way to mark the difference? We believe there is. My having evidence that you think the papers show that you are guilty is my seeing you burn them. That evidence is explained not by your being guilty but your thinking (or pretending to think) that they show you to be guilty. It is *your* evidence that you are guilty (if that is, indeed, what the papers show) that is explained by the fact that you are guilty. What is lacking between the evidence I have that you have evidence that you think shows that you are guilty and your being guilty, if you are, is what we may call a *content-connection*. If *e* is evidence that *p*, then *e* is explained by *p*'s being the case. But, as we said above, my having evidence that you have evidence that *p* is not explained by *p*. It is your evidence (if it is good evidence) that is explained by *p*. Hence my evidence that you have evidence that *p* is *not* evidence, that *p*.

We can illustrate this with an example. Suppose that Jones, a respectable mathematician, tells me that he has finally proven Goldbach's conjecture. In this case I have testimonial evidence$_1$ that Jones has evidence$_2$ that *p*, where *p* is "for any even integer *n* greater than 2, and primes *a* and *b*, $a+b=n$." Jones' evidence, if good, in effect entails *p*, since his evidence is a proof. Thus, according to (EEE), I have *eo ipso* acquired evidence$_3$ that *p*. Now, the first thing to notice is that, if by Jones' testimony I have acquired evidence at all, it is evidence different from Jones' own. For he has a proof that *p*, whereas I have only his testimony. As it is, this is not a problem for (EEE), for, as Feldman himself says, by learning that you have evidence that *p* I do not thereby "obtain your evidence."[26] The evidence I have does not entail *p*, whereas Jones' evidence does. Even if his proof is correct and I believe that it is, Jones' *telling me* that he has a proof of *p* (which is my evidence that he has one) does not entail *p* – it is the proof that does that. But then Jones' evidence that *p* is explained by *p*'s being the case, whereas the evidence I have (Jones' testimony) is explained by Jones' *belief* that he has evidence that *p*,

[26] Richard Feldman, "Evidentialism, Higher-Order Evidence, and Disagreement," 309.

hence not by something that gives me evidence that *p*.[27] He has the proof, I have his testimony.[28] He knows that *p*, I know only that he believes that *p*.

What about the fact that Jones is an authority on the matter? If I know this, is that not a reason to grant him knowledge? If I know Jones to be an expert (and honest), his testimony is clearly a reason (and a good one) for me to believe that he knows what he claims to know. However, that is not enough to say that he does. His testimony can also be explained by his mistakenly believing that he has a proof. Expertise dos not entail infallibility, hence Jones' testimony is not explained by his having, in fact, a proof. Since I have no evidence concerning whether or not he is mistaken, I have no evidence, even indirectly, about whether what he claims to know is true – even if knowing that he is an expert gives me reason to think that it is likely to be. Thus, even though having evidence that *p* gives one a reason to believe that *p*, the converse does not hold. Not everything that is a reason for believing that something is evidence that that thing is so. One can have reasons to believe (expert testimony, suspicious behaviour) even when the evidence is not, and may never be, available to one.[29] Thus even if my (supposed) peer's dissent gives me a (defeasible) reason to believe that *not-p*, this is not the same thing as having evidence that *not-p*. And if I take his dissent to be based on some evidence he has that I do not have, rather than on his assessing the evidence we both have differently, what (EEE) should be understood as saying is that evidence of evidence is a reason for at least re-visiting the evidence on which my belief is based.[30]

Finally, if we took 'evidence' to be factive, having evidence that there is evidence that *p* (whether someone has it or not) *would be* evidence that *p*. But then it would be useless as a guide to epistemic conduct in the cases – surely, most – in which my evidence that my disputant – peer or not – has evidence that *p* does not tell me that he has evidence that entails that *p*. In fact, typically the evidence

[27] Even if, as Feldman remarks, it gives me reason to believe it. The problem lies precisely in the conflation of having a reason and having evidence.

[28] This does not entail, as someone may think, that testimony is not evidence. We have agreed that I acquired evidence of *something* when Jones told me about his result. The dispute is over what that was.

[29] Compare: Sam, a reliable witness, says she has video footage locating the suspect at the crime scene. Even though I believe Sam is being honest, having thereby a reason to think the suspect guilty, if what Sam says is true, she has evidence incriminating the suspect, whereas I do not.

[30] Perhaps the distinction between direct and indirect evidence some draw is intended to mark the difference we see as one between evidence and reasons. (Harvey Siegel, "Argumentation and the Epistemology of Disagreement," 144-145, footnote 1) But, as we have just argued, something can be a reason without being even indirect evidence.

one has that the other has evidence that *p* does not even tell one whether the latter is in the least supportive of *p*. (EEE) is, surely, intended by Feldman and other advocates of it to be a fallibilist principle.

Should it be objected that there is no reason not to take (EEE) at face value and the second "E" in it to refer not just to one's disputant's belief that *p* but to the evidence on which that belief is (justifiably or not) based, we answer that that would make no difference. As long as one does not know what that evidence is (and thus, *a fortiori*, does not know whether it entails or even supports *p*) one does not have evidence that *p* by having evidence1. To have evidence that someone takes something to be evidence that *p* is not to have evidence that *p*.

We suggest that when faced with disagreement by a supposed peer or even near-peer, the rational thing to do is to take that disagreement as a reason for re-assessing the evidence on which one's belief is based and one's reasoning from that evidence.[31] (This is the kernel of truth in conciliationism.) We may dub this principle, namely, 'evidence of disagreement is reason for re-assessment,' (EDRR). Admittedly, not as catchy a slogan as (EEE) – but it has the advantage of being true.[32]

[31] We say, "supposed peer," for evidence that someone disagrees with one is, if it is evidence of anything, is evidence that the one disagreeing is not one's peer.

[32] Thanks to Harvey Siegel and Pedro Merlussi for comments on previous drafts.

A THOROUGHLY MODERN WAGER

Michael J. SHAFFER

ABSTRACT: Pascal's wager is a familiar heuristic designed to show that believing that God exists is of greater practical value than believing that God does not exist given the outcomes associated with those beliefs as understood in Christian theology. In this way Pascal argues that we that we ought to believe that God exists, independent of epistemic grounds. But, things are not easy, because he understands that belief is not subject to direct voluntary control. So, for purely practical reasons, he advises us to put ourselves in situations that will maximize our chances of acquiring the belief that God exists. In effect, he advises us to attempt to acquire that belief by indirect control. But, then the wager is not a proper decision problem since it does not involve a real choice. Additionally, there are at least two other problems that afflict the traditional wager: one involving the value of eternal damnation and one concerning the coherence of infinite utilities. In this paper the wager will be explored and a corrected version will be presented that yields a rather surprising, but theoretically correct, conclusion.

KEYWORDS: Pascal's wager, belief, acceptance, infinite utilities, decision

1. Introduction

Pascal's wager is a familiar heuristic designed to show that believing that God exists is of greater practical value than believing that God does not exist given the outcomes associated with those beliefs as understood in Christian theology. The wager has been presented in a variety of forms since its inception, but it is fundamentally based on the observation that there are only two possible factual states: God exists and God does not exist. We are to suppose also that there are two possible beliefs states we might have concerning those factual states: the belief that God exists and the belief that God does not exist. This yields four possible combinations of factual states and belief states: believing God exists and God exists, believing God exists and God does not exist, believing God does not exist and God exists, and believing God does not exist and God does not exist. The relevant outcomes that are consequences of those four states are, respectively, eternal salvation, the finite costs associated with belief in God's existence where He does not exist, eternal damnation,[1] and the finite benefits associated with

[1] See B. Pascal, *Pensées*, in *Pascal: Selections*, ed. Richard H. Popkin (New York: MacMillan, 1670/1989), 195-264. Pascal describes this outcome as "misery." As we shall see however, since it involves eternal damnation, this outcome is one that needs to be examined in greater detail.

believing God does not exist where He does not exist. On the basis of this simple looking heuristic Pascal concludes that it is better to believe that God exists than to believe that He does not exist. This is in part because the total expected value of the belief that God exists is essentially unaffected by the finite loss that one would suffer if that belief is false given the infinite value of that belief if it is true. Moreover, it is also supposed to be obvious that belief in God's existence has a greater expected value than the expected value of the belief that God does not exist. This is because whatever finite positive value that believing God does not exist has if it is true is swamped by the negative value it has if it is false (finite or not). So, according to the Wager, believing that God does not exist is clearly supposed to be the less valuable of the two beliefs in terms of total expected values. Thus understood, the wager is supposed to take the form of a standard decision-theoretic problem.

In this way Pascal argues that we that we ought to believe that God exists, independent of epistemic grounds. But, things are not so easy for two important reasons. First, there is a problem with this naïve construction of the wager because Pascal understands that belief is not subject to direct voluntary control. So, for purely practical reasons, he advises us to put ourselves in situations that will maximize our chances of acquiring the belief that God exists. In effect, he advises us to attempt to acquire that belief by indirect control. But, this suggestion raises a number of problems. For one, prior to acquiring the belief that God exists, what attitude(s) might we have with respect to the proposition that God exists? Additionally, how do those pre-belief attitudes relate to belief? Answering these questions has important implications concerning voluntarism with respect to propositional attitudes and rational commitment. In this paper it will be argued that the kind of commitment the wager involves is best modelled as a form of voluntary acceptance. This is because the kind of pre-belief attitude involved in Pascal's wager is governed by standards of pragmatic rationality that are of a different sort from those that apply to beliefs. Second, this naïve construction of the wager involves calculations of expected utility involving at least one outcome with an infinite utility. Specifically, Pascal supposes that belief in God's existence where He in fact exists has an infinite positive expected utility. But, standard decision theory is incompatible with outcomes having such utility values. As a result, here it will be suggested that in order to make sense of the wager in a thoroughly modern manner we need to introduce an alternative account of the nature of decision-theoretic rationality that allows for outcomes to have infinite values in a manner that does not raise any serious problems. Moreover, it will also be suggested that consistency demands that we treat the outcome involving

eternal damnation as having infinite negative utility for the same sorts of reasons that we attribute infinite positive utility to the outcome involving eternal salvation. Ultimately, it will be shown that recognizing all of this yields a rather surprising, modern, and elegant re-construction of the wager with a very different conclusion.

2. Preliminaries: Fixing the Traditional Wager

Typical interpretations of Pascal's wager treat it as a decision problem involving a choice about expected utility of two competing beliefs. The simplest account of this choice is understood to be one between the belief that God exists and the belief that God does not exist.[2] This choice is then supposed to be evaluated in light of the expected outcomes determined by orthodox Christian theology as they depend on the possible factual states: God exists and God does not exist. The wager so understood is supposed to involve the following elements:

O: Options $\{Bp, B\neg p\}$.
S: States $\{(\exists x)(x = G), \neg(\exists x)(x = G)\}$.
C: Outcome Values $\{V_{\infty}, V_{-\beta}, V_{\alpha}, V_{-\alpha}\}$.[3]

This is the familiar expectation form of Pascal's wager.[4] The elements in O and S do not need to be more deeply analyzed at this point, but, at this juncture, we need to be clear what the elements of C represent. Most importantly, V_{∞} is meant to represent eternal salvation and so this is an infinite positive magnitude. Why is this supposed to be the case? As Pascal understands it what is being wagered in the wager's life and, in the case of this particular outcome, "…there is here an infinity of an infinite happy life to gain."[5] So, Pascal appears to be basing this contention that the outcome involves an infinity of positive value on the idea that it is an intrinsically positive outcome that is eternal in character. In Pascal's version of the wager $V_{-\beta}$ represents the *finite* loss associated with eternal damnation. V_{α} represents the positive value associated with having those experiences precluded by orthodox Christian theological practice and so it is a

[2] See, for example, Philip Quinn, "Moral Objections to Pascalian Wagering," in *Gambling on God: Essays on Pascal's Wager*, ed. Jeff Jordan (Lanham: Rowman & Littlefield, 1994), 61-81 and Jeff Jordan, "The Many Gods Objects," in *Gambling on God: Essays on Pascal's Wager*, 101-113.

[3] An outcome value is a function on outcomes. In decision theory these are treated as utilities. So, $V_i = u(O_i)$.

[4] See Ian Hacking, "The Logic of Pascal's Wager," *American Philosophical Quarterly* 9 (1972): 186-92 for discussion of the expectation argument and the related arguments from dominance and dominating expectation found in Pascal's notes.

[5] Pascal, *Pensées*.

finite positive magnitude. $V_{-\alpha}$ represents the value lost by failing to have those experiences precluded by orthodox Christian theological practice and so it is a finite negative value. These elements are then supposed to be related in terms of the following counterfactuals: [6]

CF1: $[Bp \,\&\, (\exists x)(x = G)] \,\square\!\!\rightarrow V_{\infty}$.
CF2: $[Bp \,\&\, \neg(\exists x)(x = G)] \,\square\!\!\rightarrow V_{-\alpha}$
CF3: $[B\neg p \,\&\, \neg(\exists x)(x = G)] \,\square\!\!\rightarrow V_{\alpha}$.
CF4: $[B\neg p \,\&\, (\exists x)(x = G)] \,\square\!\!\rightarrow V_{-\beta}$.

From these counterfactuals we can generate the following decision matrix (DM1):

	Bp	$B\neg p$
$(\exists x)(x = G)$	V_{∞}	$V_{-\beta}$
$\neg(\exists x)(x = G)$	$V_{-\alpha}$	V_{α}

The expected value (EV) of an option is then defined as the sum of the expected values of the possible outcomes associated with each option. So we get the following expected values for Bp and $B\neg p$:

$EV(Bp) = V_{\infty} + V_{-\alpha}$.

$EV(B\neg p) = V_{-\beta} + V_{\alpha}$.

Since V_{∞} is an infinite magnitude the total expected value of Bp is positive and infinite and since both $V_{-\beta}$ and V_{α} are finite the expected value of $B\neg p$ will be finite whatever magnitudes those values have, $EV(Bp) > EV(B\neg p)$. So, decision-theoretical considerations are supposed to favor belief over the alternative. But all is not kosher here and we can see that there are already problems with the original wager with respect to one element in O, specifically with respect to the outcome that is supposed to represent eternal damnation.

[6] See Michael J. Shaffer, "Decision Theory, Intelligent Planning and Counterfactuals," *Minds and Machines* 19 (2009): 61-92 for extensive and critical discussion of orthodox decision theory, especially with respect to the role that counterfactuals play in decision problems.

The problem is that the very same reasons that Pascal uses to support the contention that salvation (i.e. the outcome of Bp & $(\exists x)(x = G)$) should have an infinite positive value also support the view that eternal damnation (i.e. the outcomes of B$\neg p$ & $(\exists x)(x = G)$) ought to have an infinite negative value, especially given Pascal's own commitment to orthodox Catholicism. By parity of reasoning the very eternality of eternal damnation implies that it involves an infinity of an infinitely *unhappy* life, whether or not Pascal himself acknowledges this or not.[7] So, in order then to be consistent we ought to replace the finitely valued outcome $V_{-\beta}$ with $V_{-\infty}$, an infinite negative magnitude.[8] To accommodate this insight we need to replace CF4 with CF4′:

CF4′: $[B\neg p \,\&\, (\exists x)(x = G)] \,\Box\!\!\rightarrow V_{-\infty}$.

In light of this correction we get the following decision matrix (DM2):

	Bp	B$\neg p$
$(\exists x)(x = G)$	V_{∞}	$V_{-\infty}$
$\neg(\exists x)(x = G)$	$V_{-\alpha}$	V_{α}

So we get the following expected values for Bp and B$\neg p$:

$EV(Bp) = V_{\infty} + V_{-\alpha}$.

$EV(B\neg p) = V_{-\infty} + V_{\alpha}$.

---

[7] Duff, Lycan and Schlesinger and Hacking concur on this point about assigning an infinite negative utility to damnation. See Antony Duff, "Pascal's Wager and Infinite Utilities," *Analysis* 46 (1986): 107-109, William G. Lycan and George N. Schlesinger, "You Bet Your Life: Pascal's Wager Defended," in *Reason and Responsibility*, 7th ed., ed. Joel Feinberg (Belmont: Wadsworth, 1989), 82-90 and Ian Hacking, *An Introduction to Probability and Inductive Logic*, (Cambridge: Cambridge University Press, 2001), 118.

[8] It is important to note at this point that orthodox decision theory is actually incompatible with their being outcomes with infinite values (i.e. utilities). For details see Duff, "Pascal's Wager and Infinite Utilities," 107-109, Jeff Jordan, "Pascal's Wager Revisited," *Religious Studies* (1989) 34: 419-431, Edward McClennen, "Pascal's Wager and Finite Decision Theory," in *Gambling on God: Essays on Pascal's Wager*, 115-137, Alan Hájek, "Waging War on Pascal's Wager," *Philosophical Review* 112 (2003): 27-56 and P. Bartha, "Taking Stock of Infinite Value: Pascal's Wager and Relative Utilities," *Synthese* 154 (2007): 5-52. This issue will be more fully addressed in section 6.

Since V_{∞} and $V_{-\infty}$ are infinite quantities, intuitively it would seem to be the case that they respectively swamp V_{-a} and V_a and thus wholly determine the values EV(B*p*) and EV(BØ*p*). Moreover, it is still abundantly clear from the naïve perspective that in this corrected version of the wager EV(B*p*) > EV(B¬*p*), in fact the value of B*p* is *massively* greater than that of B¬*p*. So, from a pragmatic perspective, even with this small correction it appears to be the case that we ought to adopt B*p*.

However, Pascal also famously argued that belief is not voluntary and so one cannot on this basis simply choose to believe that God exists and thus make it so.[9] In light of this recognition he argued that the best that we can do is to attempt to indirectly bring about that belief state. This indirect approach is supposed to involve things like participating in Christian practice, mingling with believers and reading Christian texts. But, this fact about our lack of direct doxastic control introduces a crucial wrinkle into the traditional wager when it is understood as a decision problem. Specifically, this characterization of the wager wrongly assumes that the options involve are subject to direct control. In standard decision theory, the options an agent has must constitute a choice for the agent and as Levi points out,

> Having a choice presupposes having options. Having the option to perform some action entails having the ability to perform the action upon choosing it. Hence, having a choice presupposes having abilities to perform various actions upon choosing them.[10]

So, the wager, as traditionally understood, *is not really a real decision problem at all*. If the traditional construal of the wager involved a choice or a decision it would have to be the case that B*p* and B¬*p* constitute possible acts subject to the direct control of the agent. This is because, if it involves a real choice or decision, then it would have to be the case that the agent has the ability to perform those acts. But, according to Pascal (and many others), these states are not subject to our direct control. So, there is no decision problem here at all. The traditional wager simply is not a well-formed decision problem.

---

[9] See Bernard Williams, "Deciding to Believe," in *Language, Belief and Metaphysics*, eds. Howard Evans Kiefer and Milton Karl Munitz (Albany: SUNY Press, 1970), 95-111 and Matthias Steup, "Doxastic Voluntarism and Epistemic Deontology," *Acta Analytica* 15 (2000): 25-56 for detailed discussion of direct doxastic voluntarism. Duff and Hacking both carefully emphasize this point in the context of Pascal's wager in, respectively, "Pascal's Wager and Infinite Utilities" and "The Logic of Pascal's Wager."

[10] Isaac Levi, *Hard Choices* (New York: Cambridge University Press, 1986), 47.

3. Belief vs. Acceptance

As a result, the actual situation in which Pascal places us is really rather different than it has been traditionally understood. The wager does not involve acts Bp and B$\neg p$ as options, because those states are not up to us (i.e. they are not directly subject to control). Hacking puts this nicely as follows:

> A decision problem requires a partition of possible actions. As Pascal sees it, you either act with indifference to God or you act in such a way that you will, in due course, believe in his existence and his edicts. There is no cant in Pascal. He accepts it as a piece of human nature that belief is catching: if you go along with pious people, give up bad habits, follow a life of 'holy water and sacraments' intended to 'stupefy one' into belief, you will become a believer. Pascal is speaking to one who is unsure whether to follow this path or whether to be indifferent to the morality of the church. The two possible acts are not 'Believe in God' and 'Do not believe.' One cannot decide to believe in God. One can decide to act so that one will very probably come to believe in God.[11]

But, Hacking does not apparently see what this actually implies about Pascal's wager. The first thing to note is that the belief states involved in the traditional construction of the wager are very much like $(\exists x)(x = \mathrm{G})$ and $\neg(\exists x)(x = \mathrm{G})$. In other words, they are better understood to be part of the set of factual states involved in the wager heuristic. What is up to us however, is whether we *accept* p or $\neg p$, and, in due course, we will explore what this entails. Nevertheless, as Pascal sees our situation, we cannot simply and directly choose to believe that God exists any more than we can choose directly that He exists, but we can commit ourselves to the proposition that He exists for prudential reasons. As he describes it, this appears to amount to *simulating* the life of a devout believer. In a moment of rhetorical flourish, Pascal describes this behavior as follows:

> You would like to attain faith and do not know the way; you would like to cure yourself of unbelief, and ask the remedy for it. Learn of those who have been bound like you, and who now stake all their possessions. These are people who know the way which you would follow, and who are cured of an ill which you would be cured. Follow the way by which they began; *by acting as if they believed, taking the holy water, having masses said, etc*. Even this will naturally make you believe, and deaden your acuteness (my italics).[12]

What is crucial to see at this point is that we can directly control this kind of commitment. One can directly, voluntarily and efficaciously choose to act as if one were a believer, even though one cannot in this way choose to be a believer.

[11] Hacking "The Logic of Pascal's Wager," 188.

[12] Pascal, *Pensées*, 259.

So, acceptance, unlike belief, is voluntary and accepting that God exists importantly involves engaging in religious activities of the familiar sort. Pascal, of course, crucially hopes that such acceptance will lead to bona fide belief and in due course we will return to discussion of the connection between such acceptance and belief. But first we need to see if the wager heuristic can be salvaged in light of this observation about the traditional wager and the nature of well-formed decision problems.

To begin, the standard propositional attitudes that are typically dealt with in epistemology and elsewhere are belief and knowledge and extant interpretations of the wager incorrectly treat it as involving only belief. But, it is also widely accepted that these are not the only propositional attitudes that can be had toward propositional contents, even if this appears to be an often forgotten or ignored point. We need only to consider the attitudes of considering *p*, grasping *p*, supposing that *p*, or of wishing that p be the case, and so on in order to see that the taxonomy of propositional attitudes is really quite diverse and complex. But, the fact that there has been relatively little discussion of these other propositional attitudes is a rather serious lacuna in philosophy, and it is likely that it has given rise to the tendency to over ascribe belief and knowledge to agents where other propositional attitudes are really at work in various cases. So, one core claim defended here is that this is just the sort of error that has afflicted traditional attempts to formally characterize the wager.

An important task then when we are considering situations or models that involve propositional commitments is to distinguish cases involving belief from those that do not involve belief. One effective way of doing this is to distinguish commitments that involve the norm of truth from those that do not involve the norm of truth. This is of course because it is widely agreed that the norm of belief is truth. By distinguishing such cases we can thereby avoid attributing inappropriate features to such situations, especially with respect to judgments of rationality. This can be effectively accomplished in the case of belief by looking at instances of the following argument scheme (*scheme 1a*):

> P1: The operant and appropriate norm in situation x involving S's attitude β toward the proposition that p is y.
>
> P2: y is not truth.
>
> P3: Truth is the norm of belief.
>
> Therefore, S's attitude β toward proposition p in situation x is not belief.

If we take seriously the claim that there are true (or even merely possible) substitution instances of this argument scheme, then it is reasonable to believe

that we can make sense of the idea that there are propositions that are believable (i.e. it is possible to believe them) and even plausible (i.e. they are not known to be false and do not seem to be false), but that are not actually believed. This is because there can be non-truth-normed rational commitments. As we shall see, some of these commitments are pragmatic in nature and so the aim of committing in those cases is broadly pragmatic, others involve commitments based on plausibility. We know, of course, that ◊B*p* does not entail B*p* as a matter of elementary modal logic, but it is also reasonable to suppose that we need not believe a proposition merely because it is plausible and rational to hold for some pragmatic reasons, or because it is merely plausible. So the upshot of this is that it is reasonable to believe that propositions can be rationally entertained but not believed, at least in the sense of plausibility or pragmatic rationality.[13] Once this possibility is seriously entertained it is apparent there are many cases of commitments that are not reasonably understood to be beliefs, but which allow us to achieve certain important and rational goals. In accordance with the recognition that many commonplace propositional commitments are not beliefs, L. J. Cohen in particular usefully distinguished belief from a particular form of acceptance.[14] He treated the latter as voluntary and pragmatically motivated, whereas the former is non-voluntary and epistemically motivated and showed how belief and acceptance have often been conflated with serious negative implications for a number of philosophical issues. Given this distinction the following argument scheme can be used to positively identify a commitment as a form of acceptance (*scheme 1b*):

> P1: If the operant and appropriate norm(s) in situation *x* involving *S*'s attitude *β* toward the proposition that *p* is plausibility and/or pragmatics, then *β* is a form of acceptance.
>
> P2: The operant and appropriate norm(s) in situation *x* involving *S*'s attitude *β* toward the proposition that *p* is plausibility and/or pragmatics.

[13] See Richard Foley, "Pragmatic Reasons for Belief," in *Gambling on God: Essays on Pascal's Wager*, 31-46 and Eddy Zemach, "Pragmatic Reasons for Belief?" *Nous* 4 (1997): 525-527 for discussion of pragmatic and epistemic justification in the context of the Wager.

[14] See L. Jonathan Cohen, *An Essay on Belief and Acceptance* (Oxford: Clarendon Press, 1992), Michael J. Shaffer, "The Privacy of Belief, Morality and Epistemic Norms," *Social Epistemology* 20 (2006): 41-54, "Three Problematic Theories of Conditional Acceptance," *Logos & Episteme* (2011): 117-125, "Doxastic Voluntarism, Epistemic Deontology and Belief-contravening Commitments," *American Philosophical Quarterly* 50 (2013): 73-82 and "Epistemic Paradox and the Logic of Acceptance," *Journal of Experimental and Theoretical Artificial Intelligence* 25 (2013): 337-353 for various discussions of acceptance and belief.

> Therefore, S's attitude β toward proposition p in situation x is a form of acceptance.

These kind of weaker but voluntary propositional commitments turn out to be quite commonplace attitudes to have toward propositions and they play roles in all sorts of behaviors like acting, exploring ideas, etc. More to the point, it will be argued here that a form of acceptance plays an important role in the proper understanding of the wager and the determination of the specific kind of acceptance that is at work in Pascal's wager is a crucial goal of this paper.

So, let us then begin by looking at the various concepts of acceptance in contrast to the concept of belief. The first important distinction to make with respect to the various attitudes of acceptance concerns the extent of such commitments. So, as we will understand it here, S's acceptance of p is *full*, if and only if S's commitment to p is governed by an appropriate closure principle.[15] A modest and reasonable version of such closure for acceptance can be simply rendered as follows:

> (JBCM) If Asp and JB$s(p \supset q)$, then Asq.

Where S's commitment is not full in this sense we will call such acceptance *limited*. The second important distinction to make among the various forms of acceptance concerns the norm that governs such cases of acceptance and thus fixes the kind of rationality that such commitments involve. So, if S's acceptance of p is *strong*, then S's commitment to p is such that p should be maximally plausible for S. Here plausibility will be understood in the following sense. S is plausible for p, if and only if, S does not know that $\neg p$ and p does not *prima facie* seem to be false to S. Where S's commitment is not strong in this sense we will call S's commitment *weak* and the norm that governs such weak forms of acceptance will be understood to be pragmatic utility. So, if S's acceptance of p is weak, then S's commitment to p is such that p should be maximally pragmatically justified for S. that Adopting the attitude of weak acceptance towards a proposition *may* involve propositions that are taken to be plausible or doing so may involve propositions that are in fact be plausible despite the agent's not taking them to be so, but

[15] Here we do not need to settle the issue about whether closure principles should be understood as involving logical or material implication, whether such closure principles should be objective rather than subjective and whether the closure principle should involve closure under belief or justified belief. So, these matters will be ignored for the purposes at hand. For further discussion of these issues see Shaffer, "Doxastic Voluntarism, Epistemic Deontology and Belief-contravening Commitments" and "Epistemic Paradox and the Logic of Acceptance." What matters here is that we understand that full acceptance involve commitment to all of the implications of an accepted proposition.

neither of these conditions are required to weakly accept a proposition. An agent might be pragmatically entertaining a proposition that happens to be plausible, but the plausibility of that proposition may not be the *rational basis* on which it is being entertained. In other words, plausibility may not be among the ultimate reasons for the adoption of that proposition. So, many such pragmatic commitments involve propositions the adoption of which is not motivated by plausibility and many commitments that aim at the adoption of plausible propositions may not be adopted for pragmatic reasons. But, where we have commitments that aim at both plausibility and pragmatic utility we have cases of what we can call *mixed acceptance* and in such cases we must be clear that the rational basis for accepting a proposition is *both* plausibility and pragmatic utility. So understood these two important distinctions yield six important categories of acceptance: strong full acceptance, weak full acceptance, strong limited acceptance, weak limited acceptance, mixed full acceptance and mixed weak acceptance. Further, more-refined versions of each of these forms of acceptance can then be determined by specifying additional features definitive of each of these types of propositional attitude. But, for the purposes at hand we can ignore these more fine-grained characterizations and focus directly on determination of which of these form(s) of acceptance are involved in the wager.

To begin, let us consider the weakest form of acceptance so understood, weak limited acceptance. As it is to be understood here, *weak limited acceptance* is a propositional attitude like belief and knowledge. Its main features are as follows:

WL1. Accepting *p* is purely voluntary.

WL2. Accepting *p* is non-evidential.

WL3. Accepting *p* is a form of supposition.

WL4. Accepting *p* is a pragmatic matter.

WL5. Accepting *p* is contextual.

WL6. Accepting *p* is not a commitment to the literal truth of *p*.

WL7. Accepting *p* is not governed by any closure principle.

More specific versions of scheme 1b arguments will then allow us to discriminate truth-normed commitments like belief from non-truth-normed commitments like this particular form of acceptance on the basis of the norm(s) it does involve. In any case, the view endorsed here is that accepting a proposition in this particular weak and limited way is a sort of voluntary, non-evidential but suppositional, pragmatic and contextual commitment that is something like

epistemically "trying out" or "using" a proposition and *some of its implications in some contexts*, and while the account of weak limited acceptance offered here shares some features in common with Cohen's account it is appreciably different because on Cohen's account acceptance is characterized by subjective closure under material implication.[16] This principle is typically understood as follows:

(SCM) If A$_s p$ and B$_s(p \supset q)$, then A$_s q$.

This closure principle is however too weak and as full acceptance is characterized here it will be understood to involve JBCM. This is simply because SCM is far too subjective in closing acceptance only under what are *believed* to be the material implication of an accepted proposition. Nevertheless, Cohen's form of acceptance is still a form of weak *full* acceptance since it does obey a form of closure. In any case, limited forms of acceptance can be distinguished from forms of full acceptance in virtue of the following general argument schemes (*schemes 2a* and *2b* respectively):

P1: In any situation x involving $\mathcal{S}$'s acceptance of p, if $\mathcal{S}$'s commitment to p is governed by some closure principle k, then that commitment is a form of full acceptance.

P2: $\mathcal{S}$'s attitude β toward p in situation C is not governed by some closure principle k.

Therefore, $\mathcal{S}$'s attitude toward proposition p in situation x is not a form of full acceptance.

P1: In any situation x involving $\mathcal{S}$'s acceptance of p, if $\mathcal{S}$'s commitment to p is governed by some closure principle k, then that commitment is a form of full acceptance.

P2: $\mathcal{S}$'s commitment to p in situation C is governed by some closure principle k.

Therefore, $\mathcal{S}$'s attitude toward proposition p in situation x is a form of full acceptance.

So, we can demonstrate that a given commitment is/is not a case of full acceptance by exploring whether an agent's acceptance satisfies some appropriate closure principle.

[16] One might also believe that such attitudes are governed by other closure principles such as closure under logical implication. Since this matter plays no role in the context of this paper, it will be ignored here. See Shaffer, "Epistemic Paradox and the Logic of Acceptance" for some discussion of the issue of closure in the context of different forms of acceptance.

We are then able distinguish cases of weak acceptance from cases of strong acceptance and from cases of mixed acceptance by determining whether they involve the requirement that S's acceptance of *p* is motivated by consideration of plausibility, whether *S*'s commitment to *p* is merely pragmatically motivated, or whether *S*'s commitment to *p* is motived both considerations of both plausibility and pragmatics. Given this distinction, *strong full acceptance* can be understood to be characterized in terms of the following principles:

SF1. Accepting *p* is purely voluntary.

SF2. Accepting *p* is non-evidential.

SF3. Accepting *p* is a form of supposition.

SF4. Accepting *p* requires that *S* takes *p* to be plausible.

SF5. Accepting *p* is not a commitment to the literal truth of *p*.

SF6. Accepting *p* is governed by JBCM.

Weak full acceptance can, similarly, be characterized as follows:

WF1. Accepting *p* is purely voluntary.

WF2. Accepting *p* is non-evidential.

WF3. Accepting *p* is a form of supposition.

WF4. Accepting *p* is a pragmatic matter.

WF5. Accepting *p* is not a commitment to the literal truth of *p*.

WF6. Accepting *p* is governed by JBCM.

Finally, *mixed full acceptance* can be characterized as follows:

MF1. Accepting *p* is purely voluntary.

MF2. Accepting *p* is non-evidential.

MF3. Accepting *p* is a form of supposition.

MF4. Accepting *p* requires that *S* takes *p* to be plausible.

MF5. Accepting *p* is a pragmatic matter.

MF6. Accepting *p* is not a commitment to the literal truth of *p*.

MF7. Accepting *p* is governed by JBCM.

Notice that in all of these cases are cases of voluntary, complete and total commitments and that the completeness and totality of these attitudes is due to the fact that they are governed by closure principles, specifically by JBCM. They

are all suppositional, non-evidential and non-truth-normed kinds of commitments and they differ only in terms of the non-evidential norms which govern them. Strong full acceptance has plausibility as a norm. Weak full acceptance has practical utility as a norm and mixed full acceptance has both plausibility and practical utility as norms.

So, let us then turn to the issue of the identifying the specific type of propositional commitment at work in the wager. Recall that, in his insightful discussion of Pascal's wager, Hacking noted the following crucial point: "The two possible acts are not 'Believe in God' and 'Do not believe.' One cannot decide to believe in God. One can decide to act so that one will very probably come to believe in God."[17] In accordance with this observation we can now establish quite easily that the commitment involved in the wager is not belief and that it is acceptance in the following manner. For the wagering agent *S*,

> P1: The operant and appropriate norms in the wager involving *S*'s attitude β toward the proposition that God exists are practical gain and/or plausibility.
>
> P2: practical gain and plausibility are not truth.
>
> P3: Truth is the norm of belief.
>
> Therefore, *S*'s attitude β toward the proposition God exists in situation the wager is not belief.

> P1: If the operant and appropriate norms in the wager involving *S*'s attitude β toward proposition that God exists in the wager is plausibility and/or pragmatics, then β is a form of acceptance.
>
> P2: The operant norm in the wager involving *S*'s attitude β toward proposition that God exists in the wager is plausibility and/or pragmatics.
>
> Therefore, *S*'s attitude β toward the proposition that God exists in the wager is a form of acceptance.

So on this basis it should be clear that the kind of commitment involved in the wager is not belief. This is because it is not motivated by a commitment to literal truth. Rather, it is aimed at some other target, and given what Pascal says about the wager the agent's options are best understood to involve a form of acceptance because that choice is motivated by pragmatic considerations. Moreover, we can also see that the wager cannot reasonably be taken to involve limited acceptance via the following consideration. If the agent's attitude involved in the wager were a form of limited acceptance, then it would not be governed by

[17] Hacking, "The Logic of Pascal's Wager," 188.

a closure principle and might also be contextual. But, this just won't do in the case of the wager. As Pascal sees it, in wagering the commitment we must have towards the proposition that God exists must be extensive enough to yield a high probability that such acceptance will lead to bona fide belief and it must be sufficiently extensive to include all of the implications of Christian practice. But, if this acceptance were limited it would allow for the agent to reject many of the implications of accepting Christian principles and practices and it would not require committing to any of them in all contexts. Thus, it seems rather unlikely that such limited acceptance of those principles would suffice to do what Pascal has in mind, the conversion of acceptance into belief by systematically feigning belief. It is simply not reasonable to believe that half-hearted, incomplete and contextually limited acceptance of those principles and practices will likely bring about bona fide belief. Given this more nuanced understanding of the propositional attitudes at work in the wager let us then return to the matter of formally characterizing the wager.

4. The Wager as a Decision Problem

The wager in all of its forms arises out of the observation that the epistemic evidence and arguments relevant to the matter of God's existence are, at best, inconclusive. In a more forceful and pessimistic frame of mind Pascal appears to believe, in fact, that they are totally ineffective and epistemically inert. For example, he says of the epistemic attempt to ground commitment to God's existence that, "Reason can decide nothing here."[18] So, the real, pragmatically motivated, wager is supposed to supplant those failed attempts to epistemically justify belief in God's existence. It does so, however, by changing the standards of rationality from epistemic rationality to a form of non-epistemic rationality. Specifically, it changes the issue from one that involves epistemic reasons to one that involves specifically pragmatic considerations. But Pascal and those who defend the wager heuristic have not appreciated all of the important implications that this entails. Since weak full acceptance and mixed full acceptance are both voluntary, governed by closure, do not have truth as a norm and do have pragmatic utility as a norm, these forms of acceptance are the only really plausible candidates for the attitudes at work in a defensible form of the wager as a decision problem. Given what Pascal says it is simply not possible that the wager involves belief because the reasons he is trying to use to motivate the disbeliever to adopt the commitment to Christian practice are pragmatic and involve a real choice.

[18] Pascal, *Pensées*, 257.

In any case, we are now in a position to see that a properly constructed version of the wager involves the following elements:

O′: Options {Ap, A$\neg p$}.

S′: States {$(\exists x)(x = G)$ & Bp, $(\exists x)(x = G)$ & $\neg$Bp, $\neg(\exists x)(x = G)$ & Bp, $\neg(\exists x)(x = G)$ & B$\neg p$}.

C′: Outcome Values {V_{∞},$V_{-\infty}$, V_{α}, $V_{-\alpha}$}.

The first thing to note about these elements that will be used to formulate a coherent construction of the wager as a decision problem is that that the elements in S′ are rather different than those used to construct the traditional wager. They are the following more complex compound factual states: God exists and the agent believes that He does exist; God exists and the agent believes that He doesn't exist; God does not exist and agent believes that He does exist and God does not exist and the agent believes that He does not exist. Secondly, the elements of O′ in this construction involve some form of full acceptance that is subject to the direct control of the wagering agent. So understood, this problem does constitute a real decision problem. It involves a real choice: the choice between committing to p or not for reasons that are pragmatic and/or related to plausibility and this is just what Pascal had in mind. In this alternate construction of the wager the elements of O′, F′ and S′ are related in terms of the following counterfactuals:

CF1: [Ap & Bp & $(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{\infty}$.

CF2: [Ap & B$\neg p$ & $(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{-\infty}$.

CF3: [Ap & Bp & $\neg(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{-\alpha}$

CF4: [Ap & B$\neg p$ & $\neg(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{-\alpha}$.

CF5: [A$\neg p$ & Bp & $(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{\infty}$.

CF6: [A$\neg p$ & B$\neg p$ & $(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{-\infty}$.

CF7: [A$\neg p$ & Bp & $\neg(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{\alpha}$.

CF8: [A$\neg p$ & B$\neg p$ & $\neg(\exists x)(x = G)$] $\Box\!\!\rightarrow V_{\alpha}$.

But, this decision problem gives rise to a very different and perhaps rather surprising decision matrix (DM3):

	Ap	A$\neg p$
Bp & $(\exists x)(x = \mathrm{G})$	V_{∞}	V_{∞}
B$\neg p$ & $(\exists x)(x = \mathrm{G})$	$V_{-\infty}$	$V_{-\infty}$
Bp & $\neg(\exists \mathrm{x})(x = \mathrm{G})$	$V_{-\alpha}$	V_{α}
B$\neg p$ & $\neg(\exists x)(x = \mathrm{G})$	$V_{-\alpha}$	V_{α}

The values assigned as the outcomes in CF1, CF3, CF4, CF6, CF7 and CF8 are straightforwardly unproblematic. But, the values assigned as the outcomes in the consequents of CF2 and CF5 are worthy of some additional commentary. The antecedent of CF2 describes the situation where the agent accepts that God exists and the agent does not believe that God exists, but God exists. In this case the agent is voluntarily committed to God's existence for practical reasons, but the agent is really in the state of disbelief about God's existence. This is the principle case that Pascal is ultimately concerned with in his discussion of the prospects for those who believe that God does not exist. However, he sees some hope here. This is because this could be a case where the disbeliever might bring about belief through acting as if God exists. The posited outcome of this state is, however, still eternal damnation, just as in the case of CF6, if such acceptance is not actually converted into true belief. This is because the disbeliever is then merely simulating in his acting as if God exists in the possible case where He exists. If the simulation is unsuccessful and does not bring about true belief, the agent still suffers eternal damnation. Such agents are ultimately not *earnest* believers and so are no better off than those who fail even to act as if God exists in the possible case where God exists. The antecedent of CF5 describes the situation where the agent accepts that God does not exist, the agent believes that God exists and God exists. In this case we have an agent who is a true believer who acts as if God did not exist in the possible situation where He exists. This could be the case of an agent who is attempting to reject God's existence by simulation of the life of a disbeliever, despite his actually believing otherwise, in much the same sort of manner that the agent who disbelieves in the case described by the antecedent of CF2 might be attempting to bring about true belief. But, since this agent is in fact a true believer he stands to gain eternal salvation unless his accepting that God does

not exist brings about true disbelief for basically the same reasons that the agent in CF2 faces eternal damnation.

However, what is most important to notice about this decision problem is that based on CF1-8 it looks like we get the following surprising expected values for A*p* and A¬*p*:

$EV(Ap) = V_{\infty} + V_{-\infty} + V_{-\alpha} + V_{-\alpha}$.

$EV(A\neg p) = V_{\infty} + V_{-\infty} + V_{\alpha} + V_{\alpha}$.

Since V_{∞} are infinite quantities $V_{-\infty}$ they cancel out and the values EV(A*p*) and EV(A¬*p*) are wholly determined by the values of the other finitary outcomes.[19] It should be clear that in this corrected construction of the wager as a well-formed decision problem EV(A¬*p*) > EV(A*p*). So, from a purely pragmatic perspective, if we entertain Pascal's invitation to wager so understood we ought to adopt A¬*p*! According to the properly constructed wager we should not accept that God exists. In other words, from the perspective of practical rationality, according to the properly reconstructed wager it is irrational to behave as if God exists.

5. From Acceptance to Belief

However, it is clear that what Pascal has in mind is that in adopting A*p* we will thereby come to adopt B*p*, or, at least, there will be very likely that accepting *p* will lead to belief that *p*. That is the unavoidable implication of his advice to the disbeliever that they should emulate believers and, "Follow the way by which they began; by acting as if they believed, taking the holy water, having masses said, etc. Even this will naturally make you believe, and deaden your acuteness."[20] Notice that if the principle that acceptance guarantees belief is included in the set of factual states that characterize the wager, then CF2 and CF4 both reduce to one of CF1 or CF3 and CF5 and CF7 will reduce to one of CF6 or CF8. Effectively, a robust enough connection between A*p* and B*p* will eliminate the act/state combinations involving A*p* and B¬*p* and A¬*p* and B*p*. If this is the case, then it would restore the result of the traditional wager because it would eliminate the outcomes of CF2 and CF4. Thereby EV(A*p*) would be changed from $V_{\infty} + V_{-\infty} + V_{-\alpha} + V_{-\alpha}$ to $V_{-\infty} + V_{-\alpha}$. Similarly, it would eliminate the outcomes of CF5 and CF7,

[19] This result also depends on being able to partition the relevant outcome values as follows: $(V_{\infty} + V_{-\infty}) + (V_{-\alpha} + V_{-\alpha})$ and $(V_{\infty} + V_{-\infty}) + (V_{\alpha} + V_{\alpha})$. This is the natural way to do the calculations however as it treats the outcomes corresponding to God exists and God does not exist as the primary basis on which to partition the outcomes in both cases.

[20] Pascal, *Pensées*, 259.

yielding a value of $V_{-\infty} + V_{\alpha}$ for EV(A¬*p*). If this is the case, then EV(A*p*) > (A¬*p*) and we get the result that it is pragmatically rational to accept that God exists in the sense that it is pragmatically rational to simulate the life of a believer. But, deriving this result is totally dependent on showing that the connection between acceptance and belief is sufficiently robust and we are now in a position to ask what we might say about the connection between A*p* and B*p* on which the traditional wager then critically hinges.[21]

As it turns out, what is specifically and crucially important for this attempt to recapture the traditional wager and its theistically inclined result is establishing that the probability of B*p* given A*p* is 1. This is easy to see based on the following consideration of the corrected version of the wager presented above. Suppose that weak/mixed full acceptance of a proposition renders belief in that proposition likely with a very high probability but not with probability 1. So the probability that an agent will believe a proposition at some later time, given that it is accepted in the weak or mixed sense at an earlier time is close to but not equal to 1 (i.e. $P(B_{t+n}p|A_tp) \approx 1$). If this is true, then (relatively speaking) the outcomes in CF2, CF4, CF5 and CF7 would be very unlikely scenarios, and CF1, CF3, CF6 and CF8 would be very likely scenarios. But, given the nature of the outcomes themselves this would not change the outcome that EV(A¬*p*) > EV(A*p*). This is because while considerations of probability can impact the expected utility values of outcomes involving finite expected values, they have no impact on the expected utility values of the outcomes involving infinite values. The infinitary nature of those magnitudes swamps any non-unitary probability no matter how close to 1 it is. For example, the value associated with CF5 is still V_{∞} even if that outcome is only infintiessimally probable due to the fact that acceptance almost always leads to belief. This is because the expected value of A¬*p* in that case is just the product of the probability that the outcome in question will come about and the magnitude of the expected value of that outcome of A¬*p* in the world state B*p* & $(\exists x)(x = G)$. So, the expected value of A¬*p* in this case is still V_{∞}. The same thing goes for all of the values and probabilities associated with CF1, CF2, CF5 and CF6. Since all of this is the case we still get the cancellation of the infinitary outcomes for both A*p* and A¬*p* in the calculation of the total expected values for those options. This makes the expected value of A*p* dependent only on the outcomes of CF3 and CF4

[21] This is not, of course, true if we were to reconstruct the wager in such a way that we replace the infinite losses and gains with vast but finite losses and gains, and this has been suggested by Jordan in "Pascal's Wager Revisited," for other reasons. In that case, the wager can be salvaged without it being the case that $P(B_{t+n}p|A_tp) = 1$, but that is not Pascal's wager. Pascal's wager clearly involves some infinitary expected values.

and the expected value of A¬p dependent only on the outcomes of CF7 and CF8. As should be clear then, it is still the case that EV(A¬p) > EV(Ap). So, unless P($B_{t+n}p$|A_tp) = 1 we still get the surprising result that the corrected and well-formed version of the wager shows that from the perspective of pragmatic rationality we should not simulate the life of the believer.

But, are there any good reasons to suppose that P($B_{t+n}p$|A_tp) = 1? There are four obvious positions one might take on the matter, and they involve treating the connection between Ap and Bp as (1) a strong modal connection, (2) a matter of natural law, (3) a logical implication, or (4) a brute unitary conditional probability. Let us begin by considering (1). First, is it reasonable to suppose that there is a strong modal tie between accepting and believing that would entail that P($B_{t+n}p$|A_tp) = 1? Clearly, if □(A_tp ⊃ $B_{t+n}p$), then P($B_{t+n}p$|A_tp) = 1 = 1. But, □(A_tp ⊃ $B_{t+n}p$) seems simply to be false. It is far too strong to even be remotely plausible. There is nothing at all impossible about the existence of cases where an agent has the following attitudes: A_tp & B_{t+n}¬p. They are simply cases where simulating belief in p for practical reasons does not successfully result in later believing that p. There is nothing at all contradictory about such cases. Moreover, surely part of the gravity that Pascal attaches to the wager is that it is no sure thing that this kind of acceptance will lead to belief, certainly not as a matter of alethic necessity. He seems to be acutely aware that actual failure in this regard is a real possibility and that our efforts at simulation of belief thus require work and earnest hope that our efforts to bring about belief are successful. Otherwise we face eternal damnation. Second, is it reasonable to believe that there is a strong nomological tie between accepting and believing that would entail that P($B_{t+n}p$|A_tp) = 1? To this end, suppose that the necessitarian view of laws of nature is correct and that N(A_tp, $B_{t+n}p$) is true (i.e. that acceptance nomologically necessitates belief).[22] If this were true then, it would be the case that P($B_{t+n}p$|A_tp) = 1 in the actual world and in those possible worlds characterized by the same laws. But, is it reasonable to suppose that an agent's having the attitudes A_tp & B_{t+n}¬p is nomologically impossible? Surely it is not, and it is simply not reasonable to suppose that cases involving A_tp & B_{t+n}¬p are precluded by the laws of nature in the actual world or in close possible worlds characterized by the same laws. This is simply because there are, in fact, actual cases where simulating belief in p does not successfully result in later believing that p. Consider for example, any number of cases

[22] For elaboration of the necessitarian view see David M. Armstrong, *What is a Law of Nature?* (Cambridge: Cambridge University Press, 1983), Fred Dretske, "Laws of Nature," *Philosophy of Science* 44 (1997): 248-268 and Michael Tooley, "The Nature of Laws," *Canadian Journal of Philosophy* 7 (1977): 667-698.

involving acting, or pretending, or indoctrinating where the agent accepts a set of propositions and their implications for some period of time but does not ultimately come to believe them. So, it is not reasonable to suppose that acceptance will lead to belief as a matter of nomological necessity and there is no reason to believe that $P(B_{t+n}p|A_tp) = 1$ is true on that basis. Suppose then that one were to adopt the yet weaker view that $Ap \supset Bp$. If this implication were true at the actual world, as a matter of mere regularity, then it would also be the case that $P(B_{t+n}p|A_tp) = 1$. But, again, this is totally implausible because there are clearly actual cases where simulating belief in a proposition does not lead to belief in that proposition at a later time. So, this suggestion fares no better than the two stronger alternatives we have considered. Finally, let us consider $P(B_{t+n}p|A_tp) = 1$ itself. Is there any good reason to suppose that the key proposition about the probabilistic relationship between Ap and Bp is itself true? Certainly the answer is no and this claim is not true for exactly the same sorts of reasons we have just examined in the context of stronger attempts to yield that result. It simply isn't true that acceptance always leads to subsequent belief. $P(B_{t+n}p|A_tp) \neq 1$. The real problem with the wager then, however, is that, given any interpretation of the connection between acceptance and belief, when the wager is properly rendered as a decision problem involving voluntary acceptance it favors $A\neg p$ as a matter of pragmatic rationality. This is because there is no plausible way to justify the claim that $P(B_{t+n}p|A_tp) = 1$ and this is necessary for recapturing the result of the original but ill-formed version of the wager. As a result, when the wager is properly constructed as a well-formed decision problem involving acceptance rather than belief $EV(A\neg p) > EV(Ap)$. So, if pragmatic considerations are all we have to go on, then we should not accept Christian practice. We should behave as if the proposition that God exists is false. This is what prudence actually advises if this construction of the wager is theoretically sound.

6. Infinite Utilities, Maximin and the Modernized Wager

However, there is still one deeply serious problem with wager arguments that must be contended with. Specifically, as mentioned in section 1, standard decision theory is notoriously incompatible with the idea that there can be outcomes with infinite valued utilities.[23] This renders the results of the original wager and the modernized re-construction moot. None of these decision problems can be framed in terms of the standard theory of utility and this looks to be essential to these

[23] See Duff, "Pascal's Wager and Infinite Utilities," Jordan, "Pascal's Wager Revisited, McClennen, "Pascal's Wager and Finite Decision Theory," Hájek, "Waging War on Pascal's Wager," and Bartha, "Taking Stock of Infinite Value."

sorts of arguments. Absent some way to incorporate infinite utilities into decision theory, we simply cannot meaningfully apply the principle of maximizing expected utility in these problems. A simple solution to this problem will be introduced here that appeals to the minimax principle, but it still allows for infinite utilities and corresponding preferences. It avoids this problem by completely avoiding appeals to probabilities in the argument.

In order to set the stage for the first possible solution to the problem of infinite utilities let us examine why the standard theory of utility involved in orthodox decision theory rules out outcomes with such utilities. Standard utility theory is based on the idea that if an agent's preferences obey a certain set of axioms, then they can be represented as a utility function that exhibits certain supposedly desirable algebraic features. These axioms are introduced on the basis of their supposed intuitive (i.e. a priori) plausibility. Let "$x \preccurlyeq y$" mean "x is weakly preferred to y", "$x \prec y$" mean "x is strictly preferred to y" (i.e. x is weakly preferable to y but x is not indifferent relative to y) and "$x \backsim y$" mean "x is indifferent relative to y" (i.e. x is weakly preferred to y and y is weakly preferred to x). Let O_i, O_j and O_k represent distinct outcomes and $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$,… represent distinct probability values. Finally, let $u(O_i)$ be a function representing a real numbered valuation of O_i. Given these basic representations we can then represent a gamble with a probability p of winning O_1 and a probability q of winning O_2 as [$\mathbf{p}O_1$, (1 – $\mathbf{p}$) O_2]. In terms of these representations, the axioms are used to characterize what is intuitively taken to be rational preference orderings are as follows.[24] First we have the ordering axiom:

(U1) The preference relation $\succcurlyeq$ is a total ordering that is reflexive and transitive.

Second, we have the better prizes axiom:

(U2) For a fixed probability, prefer the gamble with a greater prize.

Third, we have the better chances axiom:

(U3) For a fixed prize prefer the gamble with a greater probability.

Fourth, we have the reduction of compound gambles axiom:

(U4) Compound gambles are to be evaluated in terms of the probability calculus.

Finally, we have the Archimedean or Continuity axiom:

[24] This is the standard presentation of this representation theorem and it closely follows Bartha, "Taking Stock of Infinite Value." See Michael D. Resnik, *Choices* (Minneapolis: University of Minneapolis Press, 1987) and Gerald Gaus, *On Philosophy, Politics and Economics* (Belmont: Wadsworth, 2008) as well.

(U5) For any outcome that is ranked between two others there is a gamble between the more preferred and less preferred outcomes such that the agent is indifferent between it and the outcome ranked in between the more preferred and less preferred outcomes.

Formally, in terms of $\preccurlyeq$ these axioms can be presented as follows:

(U1.0) For any O_i and O_j either $O_i \preccurlyeq O_j$ or $O_j \preccurlyeq O_i$,

(U1.1) For any O_i, $O_i \preccurlyeq O_i$.

(U1.2) For any O_i, O_j and O_k, if $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then $O_i \preccurlyeq O_k$.

(U2) $O_i \preccurlyeq O_j$, iff, for any $0 \leq \mathbf{p} \leq 1$ and any O_k, $[\mathbf{p}O_k, (1 - \mathbf{p})O_i] \preccurlyeq [\mathbf{p}O_k, (1 - \mathbf{p})O_j]$ and $[\mathbf{p}O_i, (1 - \mathbf{p})O_k] \preccurlyeq [\mathbf{p}O_j, (1 - \mathbf{p})O_k]$.

(U3) If $O_i \preccurlyeq O_j$, then for any $0 \leq \mathbf{p}, \mathbf{q} \leq 1$, $\mathbf{p} \geq \mathbf{q}$ iff $[\mathbf{p}O_i, (1 - \mathbf{p})O_j] \preccurlyeq [\mathbf{q}O_i, (1 - \mathbf{q})O_j]$

(U4) For any O_i and O_j and $\mathbf{p}$, $\mathbf{q}$, $\mathbf{r}$ such that $0 \leq \mathbf{p}, \mathbf{q}, \mathbf{r} \leq 1$, $[\mathbf{p}[\mathbf{q}O_i, (1 - \mathbf{q})O_j], (1 - \mathbf{p})[rO_i, (1 - \mathbf{r})O_j]] \backsim [\mathbf{r}O_i, (1 - \mathbf{t})O_j]$ for $\mathbf{t} = \mathbf{pq} + (1 - \mathbf{p})\mathbf{r}$.

(U5) If $O_i \preccurlyeq O_j$ and $O_j \preccurlyeq O_k$, then there is a $\mathbf{p}$ such that $0 \leq \mathbf{p} \leq 1$ and $O_j \backsim [\mathbf{p}O_i, (1 - \mathbf{p})O_k]$.

If an agent's preferences satisfy these axioms then those preferences can be represented by a real valued utility function $u(O_i)$ obeying the following two important conditions:

(C1) $O_i \preccurlyeq O_j$ iff $u(O_i) \leq u(O_j)$.

(C2) $u([\mathbf{p}O_i, (1 - \mathbf{p})O_j]) = \mathbf{p}u(Oi) + (1 - \mathbf{p})u(O_j)$.

The Expected Utility Theorem, the core idea behind utility theory, is then simply this claim that if one's preferences satisfy U1-U5, then those preferences can be represented as a real valued utility function satisfying C1 and C2.[25] In other words, formal utilities are a real-valued measure of preference and the value V_i of an outcome O_i is just $u(O_i)$. What is key here is that U1-U5 implicitly rule out infinite utilities and thus rule out a priori that agents can have corresponding

[25] This is just the standard way of introducing utility theory via a representation theorem. This approach takes it as given a priori that U1-U5 are true. Recently, this approach to legitimizing decision theory has been challenged in Kenny Easwaran, "Decision Theory without Representation Theorems," *Philosophers' Imprint* 14 (2014): 1-30 and by Christopher J. G. Meacham, C. and Jonathan Weisberg, "Representation Theorems and the Foundations of Decision Theory," *Australasian Journal of Philosophy* 89 (2011): 641-663.

preferences. Specifically, U2, U3 and U5 are incompatible with there being infinitely valued outcomes.[26]

As a result, a simple way to address the problem of infinite utilities in the modernized wager involves treating that wager as a decision under total ignorance/uncertainty involving infinite utilities. In such decision situations it is acknowledged that no probabilities can be meaningfully assigned to the outcomes and so the advice about what to do in such situations is wholly a function of the utilities involved. Given this approach we simply acknowledge that there are no probabilities that can be meaningfully assigned in the wager and so there are no expected utilities defined as products of probabilities and utilities involved in the wager. As we have seen this comports well, however, with Pascal's own understanding of the problem about which he makes the following claim: "Reason can decide nothing here."[27] As a result, the standard rule of maximizing expected utility does not apply. Rather, in cases where the potential loses are great and where we have no information about probabilities other than that the probabilities of all the outcomes are non-zero many decision theorists suggest that we use the maximin rule to determine what to do.[28] This has some additional appeal to it as well given Pascal's comments about our lack of epistemic reasons that pertain to the question of God's existence that were examined earlier and which can be usefully extrapolated to the modernized wager. If this is the case, then as long as the outcomes associated with [Ap & Bp & ($\exists x$)(x = G)] and [Ap & B$\neg p$ & ($\exists x$)(x = G)] and with [A$\neg p$ & Bp & ($\exists x$)(x = G)] and [A$\neg p$ & B$\neg p$ & ($\exists x$)(x = G)] are finite and symmetric the necessary cancellation occurs and the verdict that we should reject the life of the believer holds. This is because according to the maximin rule we are to maximize the minimum. So we look at the decision table and look at the worst outcomes for the two acts A$\neg p$ and Ap. It turns out that this is the case for A$\neg p$ where B$\neg p$ & ($\exists x$)(x = G) and this is the case for Ap also where B$\neg p$ & ($\exists x$)(x = G). But these maximal minima are equal (i.e. $-\infty$). So, according to the lexical maximin rule we are to look at the next lowest outcome(s) of A$\neg p$ and Ap. In the case of Ap we have the next lowest minima where Bp & $\neg$($\exists x$)(x = G) and where B$\neg p$ & $\neg$($\exists x$)(x = G). This value is $-\alpha$ in both cases. In the case of A$\neg p$ the next lowest minima are where we have Bp & $\neg$($\exists x$)(x = G) and B$\neg p$. The value

[26] See McClennen, "Pascal's Wager and Finite Decision Theory" and Bartha, "Taking Stock of Infinite Value" for details.

[27] Pascal, *Pensées*, 257.

[28] See Abraham Wald "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *The Annals of Mathematics* 10 (1939): 299-326, "Statistical Decision Functions Which Minimize the Maximum Risk," *The Annals of Mathematics,* 46 (1945): 265-280 and Resnik, *Choices*.

in both cases is α. So, the lexical maximin rule tells us to do A$\neg p$ in the case where we cannot assign probabilities to the states of the world involved and the surprising verdict of the modernized reconstructed wager still holds.[29]

[29] There is also a related, simple and obvious way to yield the same result in terms of a simple dominance argument, if one objects to the maximin argument offered here. An act A dominates an act B if for every outcome the utility of A is equal to or greater than the utility of B and for at least one outcome the utility of A is greater than that of B. The dominance rule, then says something like, where probabilities cannot be meaningfully assigned, do the dominant act. The corrected version of the wager presented here then suggests two arguments in favor of the non-acceptance conclusion.

DISCUSSION NOTES/ DEBATE

PRAGMATIC OR PASCALIAN ENCROACHMENT? A PROBLEM FOR SCHROEDER'S EXPLANATION OF PRAGMATIC ENCROACHMENT

Andy MUELLER

ABSTRACT: I argue against Schroeder's explanation of pragmatic encroachment on knowledge. In section 1, I introduce pragmatic encroachment and point out that an explanation of it should avoid Pascalian considerations. In section 2, summarize the key aspects of Schroeder's explanation of pragmatic encroachment. In section 3, I argue that Schroeder's explanation faces a dilemma: it either allows for an objectionable form of Pascalian encroachment or it fails to be a fully general explanation of pragmatic encroachment.

KEYWORDS: knowledge, pragmatic encroachment, reasons to withhold

1. Introduction of Pragmatic Encroachment

The following case pair has spurred a number of debates:

> **Low Stakes**: Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. It is not important that they do so, as they have no impending bills. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Realizing that it isn't very important that their paychecks are deposited right away, Hannah says, 'I know the bank will be open tomorrow, since I was there just two weeks ago on Saturday morning. So we can deposit our paychecks tomorrow morning.'
>
> **High Stakes**: Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their paychecks by Saturday. Hannah notes that she was at the bank two weeks before on a Saturday morning, and it was open. But, as Sarah points out, banks do change their hours. Hannah says, 'I guess you're right. I don't know that the bank will be open tomorrow.'[1]

[1] This variation of the bank cases is taken from: Jason Stanley, *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).

The puzzling thing about these cases is that Hannah seems to know in Low Stakes, but fails to know in High Stakes, although the cases differ only in what is at stake for her. Traditional theories of knowledge deny that what is at stake affects whether a true belief amounts to knowledge—only truth-conducive factors can make a difference. One reaction to the cases, albeit surely not the only one, is that traditional theories are wrong. Some take the cases to imply that pragmatic encroachment on knowledge, henceforth (PE), is true: whether a true belief amounts to knowledge does not only depend on truth-conducive factors, but also on practical factors, e.g. what is at stake.[2]

(PE) is a controversial thesis. Mark Schroeder[3] surmises that this is due in part to the following dictum. It is consensus among epistemologists that Pascalian considerations (i.e. the benefits of having a belief) do not affect whether a true belief amounts to knowledge. But many view Pascalian considerations as paradigmatic for practical factors. Since the stakes are also a practical factor, many will be suspicious of their influence on knowledge.

Schroeder's ambition is not to argue that (PE) is true, as many others do[4], but to offer an explanation of how it could be true. It is this explanation that is the target of my criticism, not (PE) itself. I think that an explanation of (PE) should respect the consensus that Pascalian considerations have no place in a theory of knowledge. Thus an explanation of (PE) that allows Pascalian considerations to play a role so that pragmatic encroachment turns into Pascalian encroachment ought to be rejected. Schroeder seems to accept this condition for a proper explanation.[5] In the following, I will investigate whether his explanation fulfills this condition.

---

[2] Among the main proponents of pragmatic encroachment on knowledge are Stanley *Knowledg*e, Jeremy Fantl and Matthew McGrath, *Knowledge in an Uncertain World* (Oxford: Oxford University Press, 2009), Brian Weatherson "Knowledge, Bets, and Interests," in *Knowledge Acriptions*, ed. Jessica Brown and Mikkel Gerken (Oxford: Oxford University Press, 2012), Jacob Ross and Mark Schroeder, "Belief, Credence, and Pragmatic Encroachment," *Philosophy and Phenomenological Research* 88 (2014): 259-288. It is at least entertained in John Hawthorne, *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).

[3] Mark Schroeder, "Stakes, Withholding and Pragmatic Encroachment on Knowledge," *Philosophical Studies* 160 (2012): 266.

[4] See footnote 2 for a list of works arguing in favor of (PE) without necessarily giving an explanation of how (PE) works.

[5] Schroeder, "Stakes, Withholding," 282.

2. Schroeder's Explanation of Pragmatic Encroachment

The general idea behind Schroeder's explanation of (PE) is this. In High Stakes, Hannah fails to know that *p* because it is not rational for her to believe that *p*.[6] Despite her having evidence for *p*, it can still be irrational to believe that *p*, because there can be reasons to withhold believing that outweigh the reasons for belief provided by the evidence. These reasons to withhold are not merely additional evidence, as the evidence in both cases seems to remain constant. The high stakes, which are a practical factor and which are the only difference between the cases, could be conceived as providing Hannah with a reason to withhold. These reasons to withhold are offered as an explanation of the shift in knowledge throughout the cases and since they are connected to a practical factor, this is also an explanation of how (PE) could be true.

To assess this explanation, we must get clear on Schroeder's conception of reasons to withhold. Reasons to withhold on *p* are reasons to not make up one's mind about *p*. One natural suggestion is that any disadvantage of forming a belief is a potential reason to withhold. Among the disadvantages of forming a belief, Schroeder sees the costs of error, which are central to his account of reasons to withhold. Schroeder identifies two types of error. Type-1 error consists in forming a belief in a falsehood. Type-2 error consists in withholding and thereby missing out on having a true belief. Reasons to withhold are then derived from the preponderance of the costs of type-1 error over the costs of type-2 error. In other words, if it is costlier to have a false belief than to miss out on having a true belief, one has a reason to withhold. Schroeder holds that withholding belief is more rational than believing when the costs of type-1 error exceed the costs of type-2 error and also outweigh the evidence.[7]

Schroeder is upfront about his talk of outweighing or comparing costs being an idealization. Nonetheless, we can appreciate how his account intends to handle the bank cases. In Low Stakes, the costs of Type-1 error are very low, as are the costs of type-2 error. Nothing serious happens if Hannah's belief turns out to be false and there are no serious consequences if Hannah fails to believe that the bank is open on Saturday. She will just have to stand in line on Friday. So the costs of Type-1 error do not exceed the costs of Type-2 error. Therefore, and given Hannah's evidence, it is rational to believe instead of to withhold. In High Stakes, the costs of Type-1 error are high. If Hannah's belief turns out to be false, she will be late on the important payment. The costs of Type-2 error are very low. If

[6] Schroeder, "Stakes, Withholding," 268.

[7] Schroeder, "Stakes, Withholding," 281.

Hannah does not believe that the bank is open on Saturday, she will not act on this proposition and she will have to endure the small annoyance of standing in line on Friday. In High Stakes, the costs of Type-1 error clearly exceed the costs of Type-2 error. Therefore, even given Hannah's evidence, she has stronger reasons to withhold believing and that is why a belief that the bank is open on Saturday would fail to be epistemically rational and why Hannah fails to know.

This seems to be an appealing explanation of (PE) that respects the ban of Pascalian considerations. The explanatory work seems to be done by costs of certain errors. It is not the benefits of having a belief that accounts for the difference in knowledge between Low Stakes and High Stakes. Unfortunately, I think this appearance is misleading.

3. A Problem for Schroeder's Explanation of Pragmatic Encroachment

I will now argue that Schroeder's explanation of (PE) faces a dilemma: either it inadvertently allows for Pascalian encroachment or it fails to be a fully general explanation of pragmatic encroachment. The problem arises due to the following case that Schroeder himself gives:

> **Forced Choice**: Hannah and her wife Sarah are out driving on Saturday morning, at twenty minutes to noon. Since they have an impending bill coming due, and very little in their account, it is very important that they deposit their pay-checks that day, but they have so far forgotten to do so. Sarah remembers that they still haven't deposited their paychecks from Friday, but points out that just one of their bank's two branches is open until noon on Saturdays, but she can't remember which, and there is only time to try one. Hannah says, 'Oh, I remember being at the branch on Chapala Street two weeks ago on Saturday. It's the one that is open today.' Hannah is right—the branch on Chapala Street is the one that is open on Saturday.[8]

The significant detail in this case is that Hannah cannot engage in further inquiry about the hours of the bank before she makes a decision and that she is forced into deciding to go to one of the banks. Schroeder does not explicitly say that Hannah knows in this case, nor does he explicitly deny it.[9] As I will argue now, the verdict that his explanation of (PE) obliges him to hold is that Hannah in Forced Choice knows.

[8] Schroeder, "Stakes, Withholding," 278.

[9] Footnote 12 in Schroeder "Stakes, Withholding" points to Schaffer, "The Irrelevance of the Subject: Against Subject Sensitive Invariantism," *Philosophical Studies* 127 (2006): 87-107. Schaffer gives a similar case and thinks that the subject in this case knows. This might indicate that Schroeder agrees with Schaffer.

Schroeder seems bound to hold that Hannah does not have a reason to withhold. The costs of Type-1 error are high, as high as in High Stakes. But unlike in High Stakes, the costs of Type-2 error are also high in Forced Choice. If Hannah fails to form a belief at all and is therefore unable to make a decision, this would guarantee that the worst possible outcome obtains. Thus in Forced Choice, there is no preponderance of the costs of Type-1 error over Type 2 error, and consequently no longer a reason to withhold. Additionally, it seems entirely rational for Hannah to form a belief. She has some evidence, and since no more can be acquired, she is rational in believing that the bank on Chapala street is open. By making up her mind and forming the belief, which enables her to make a decision, Hannah at least stands a chance to avoid disaster.

But if Schroeder is committed to this, then he is committed to hold that Hannah knows in Forced Choice. If the presence of a reason to withhold was what caused Hannah's lack of knowledge in High Stakes, then the absence of such a reason to withhold should make it the case that Hannah knows in Forced Choice. Likewise, while it is rational to withhold in High Stakes, it does not seem rational to withhold in Forced Choice. But then one cannot say that a lack of epistemic rationality in believing causes Hannah's lack of knowledge. Since we have now exhausted the resources of Schroeder's explanation, it seems that he is bound to hold that Hannah knows in Forced Choice.

I think this is the wrong result. My own intuition is that Hannah does not know in Forced Choice. But I will not insist on this intuition. My point is not merely that Schroeder's explanation of (PE) leads to a counterintuitive result in Forced Choice. More importantly, Forced Choice brings out that Schroeder's explanation of (PE) allows for Pascalian encroachment.

Schroeder characterizes reasons to withhold as reasons not to make up one's mind. It seems natural that these reasons to withhold should also be sensitive to the costs of not making up one's mind, that is Type-2 error. But the costs of not making up one's mind should not be a knowledge making feature. The costs of not making up one's mind are determined by the benefits of making up one's mind. The costs of not making up one's mind on the existence of God are determined by the benefits making up one's mind on the existence of God. We should now see that something has gone wrong. It seems that costs of Type-2 error are closely tied to Pascalian considerations—the benefits of forming a belief.

Schroeder's explanation of (PE) allows Pascalian considerations to enter into epistemology. The benefits of believing should not be a knowledge making feature. But this is what they could be if we consequently apply Schroeder's notion of reasons to withhold to Forced Choice. For Hannah, it is clearly

beneficial to form a belief in Forced choice, as the costs of Type 2 error are very high. By making up her mind, she at least has a chance of making it to the right bank in time. But since the costs of Type-2 error are tied to how beneficial forming a belief in a situation is, they are tied to traditional Pascalian considerations. But if the costs of Type-2 error affect whether one has reasons to withhold and thus whether one knows, then, at least in cases like Forced Choice, it turns out that Pascalian considerations are a knowledge-making feature. They are, because the only difference between Hannah in High Stakes and Forced Choice and Low Stakes is the presence of a reason to withhold. If the absence of a reason to withhold makes it that Hannah knows in Low Stakes, then it also does in Forced Choice. Since Schroeder's explanation of (PE) allows for Pascalian encroachment, we should reject this explanation as it fails an important condition for a proper explanation, as was set out in section 1.

There are at least two responses available to Schroeder. He might want to rid himself of Type-2 error and make reasons to withhold entirely dependent on Type-1 error. While this gets around the problem of Pascalian encroachment, this still leads to questionable results in Forced Choice. The costs of Type-1 error in Forced Choice and in High Stakes are equally high. If the costs of Type-1 error provide reasons to withhold, then they should do so in both cases. But they should also be equally strong in both cases and make it rational for Hannah to withhold. While there is nothing objectionable in saying that Hannah's reason to withhold makes it rational to withhold in High Stakes, this is not true in Forced Choice. Clearly, Hannah would be irrational in withholding, as she would then be guaranteed disastrous consequences because she would fail to make a choice since she lacks the relevant belief. This shows that Schroeder cannot just modify his account of reasons to withhold to incorporate just costs of Type-1 error in order to avoid the problem of Pascalian encroachment.

Moreover, if we assume that Hannah does not know in Forced Choice, we see that there is something amiss with the strategy to explain (PE) through reasons to withhold and a lack of epistemic rationality in believing. There seems to be no good reason to withhold in Forced Choice and, at least to me, it seems that in Forced Choice, Hannah would be rational in believing that the bank on Chapada street is open, as her memory provides her with some evidence. But still Forced Choice seems to be a case in which Hannah fails to know, like in High Stakes. This suggests that Schroeder's explanation lacks in generality to account for all relevant cases.

Of course, Schroeder is free to hold that the lack of knowledge in Forced Choice can be explained by other means, which is the second possible response to

my challenge. While this certainly puts pressure on the general idea that (PE) can be explained by reasons for withholding, it is a way to defend this idea against my charge of Pascalian encroachment. Perhaps Schroeder is willing to explore this route.

While it would be premature to call the case settled, we can summarize that Forced Choice raises the following dilemma for Schroeder: if he maintains his account for reasons to withhold, then this would suggest that Hannah knows in Forced Choice. As I have pointed out, this would mean that Schroeder's explanation of (PE) is committed to Pascalian encroachment. If Schroeder wants to agree that Hannah does not know in Forced Choice, then his explanation of (PE) lacks in generality. As I have pointed out, if Hannah fails to know in Forced Choice, then this seems not to be caused by reasons to withhold or by a lack of epistemically rational belief. In closing, I want to state clearly once more that I do not think that my arguments speak directly against (PE). However, they do suggest that we are lacking a proper explanation of how (PE) could work, as to the best of my knowledge, Schroeder's explanation is the only one currently on offer.

STILL STUCK ON THE BACKWARD CLOCK: A REJOINDER TO ADAMS, BARKER AND CLARKE

John N. WILLIAMS

ABSTRACT: Neil Sinhababu and I presented *Backward Clock*, an original counterexample to Robert Nozick's truth-tracking analysis of propositional knowledge. In their latest defence of the truth-tracking theories, "Methods Matter: Beating the Backward Clock," Fred Adams, John A. Barker and Murray Clarke try again to defend Nozick's and Fred Dretske's early analysis of propositional knowledge against *Backward Clock*. They allege failure of truth-adherence, mistakes on my part about methods, and appeal to charity, 'equivocation,' reliable methods and unfair internalism. I argue that these objections all fail. They are still stuck with the fact that the tracking theories fall to *Backward Clock*, an even more useful test case for other analyses of knowledge than might have first appeared.

KEYWORDS: sensitive belief, extra-sensitive belief, sensitive methods, truth-adherence, reliable methods, Backward Clock

In a seminal paper, "Resurrecting the Tracking Theories," Fred Adams and Murray Clarke argued persuasively and ingeniously that a number of well-known examples that appear clearly fatal to Robert Nozick and Fred Dretske's truth-tracking analyses of knowledge are not really counterexamples at all.[1] These include Ray Martin's *Racetrack*, George Pappas and Marshall Swain's *Generator*, and Laurence Bonjour's *Clairvoyant*,[2] as well as Saul Kripke's *Red Barn*, his *Deceased Dictator*, and his *Sloppy Scientist*.[3] So taken was I with this defence that I assumed that the truth-tracking analyses were impregnable. Until that is, Neil

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

[2] See Raymond Martin, "Empirically Conclusive Reasons and Scepticism," *Philosophical Studies* 28, 3 (1975): 215-217; "Tracking Nozick's Sceptic: A Better Method," *Analysis* 43, 1 (1983): 28-33; George S. Pappas and Marshall Swain, "Some Conclusive Reasons against 'Conclusive Reasons'," *Australasian Journal of Philosophy* 51, 1 (1973): 72-76; and Laurence Bonjour, "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy* 5, 1 (1980): 53–74.

[3] Kripke gave these widely known examples at a session of the American Philosophical Association in the 1980s.

Sinhababu and I hit upon a counterexample with a special feature. In "The Backward Clock, Truth-Tracking, and Safety," we presented *Backward Clock*, an original counterexample to Robert Nozick's truth-tracking analysis of propositional knowledge.[4] We showed that analyzing knowledge in terms of three formulations of safe belief cannot withstand *Backward Clock* either. In their reply, "Beat the (Backward) Clock," Fred Adams, John A. Barker and Murray Clarke gave reasons why *Backward Clock* is not a counterexample.[5] In my rejoinder, "There's Nothing to Beat a Backward Clock: A Rejoinder to Adams, Barker and Clarke," I argued that these reasons fail and argued that *Backward Clock* shows that Dretske's early analysis of knowledge is too weak as well.[6] In their latest defence of the truth-tracking theories, "Methods Matter: Beating the Backward Clock," Adams *et al* try again to extricate themselves from the failure of the truth-tracking analyses.[7] I believe that this attempt fails. They are still stuck with the *Backward Clock* because they have got stuck on it. As Shakespeare might have said of it, "nothing 'gainst Time's scythe can make defence."[8]

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

Nozick's analysis of propositional knowledge is as follows.

> *S* knows that *p*, using method *M* of arriving at a belief whether *p*, just in case
>
> (1) *p*.
>
> (2) *S* believes, using *M*, that *p*.
>
> (3) In the closest (that is, most similar) worlds to the actual world in which not *p* (and in which *S* uses *M*), *S* does not believe that *p*.
>
> (4) In the closest (that is, most similar) worlds to the actual world in which *p* (and in which *S* uses *M*), *S* believes that *p*.[9]

---

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

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

[6] John N. Williams, "There's Nothing to Beat a Backward Clock: A Rejoinder to Adams, Barker and Clarke," *Logos & Episteme* VII, 3 (2016): 363-378.

[7] Murray Clarke, Fred Adams, and John A. Barker, "Methods Matter: Beating the Backward Clock", *Logos & Episteme* VIII, 1 (2017): 99-112.

[8] Sonnet 12.

[9] This formulation is faithful to Nozick, although it is not *verbatim*. In *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 179, he says the following.

> Let us define a technical locution, *S* knows, via method (or way of believing) *M*, that *p*.

(3) is commonly known as the 'sensitivity condition,' meaning that *S*'s belief that *p* is sensitive to falsehood; roughly, she would not have that belief if it were false. (4) is commonly known as the 'adherence condition,' meaning that *S*'s belief that *p* adheres to the truth; roughly, were she to have that belief in slightly changed circumstances, then it would still be true. A belief that is both sensitive to falsehood and adherent to truth is said to be 'truth-tracking.'

In my rejoinder to Adams *et al*, I showed that *S* does not know that *p* in *Backward Clock,* but that this example satisfies (1)-(4), thus showing that Nozick's analysis, as given above, is too weak, predicting knowledge where there is ignorance. In order to support this claim, I first gave two other examples, *Normal Clock* and *Stopped Clock*.[10] I *originally* described *Normal Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because the clock has continued to work perfectly reliably.[11]

In their latest defence, Adams *et al* countenance this as claiming that your method is "looking at the clock and determining what it says."[12] As I said in my rejoinder, this is not what I stipulated.[13] Your method of ascertaining the time you wake is to observe, during the period from 4:00 pm to 5:00 pm (since that is the period during which you nap, not knowing when you will wake) the position of its hands. To try to make this perfectly clear, here is my *current* formulation of *Normal Clock*.

> You habitually nap between 4:00 pm and 5:00 pm. Your method of ascertaining the time you wake is to observe, between 4:00 pm and 5:00 pm, the position of

---

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

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

[10] Williams, "Nothing to Beat," 365-366.

[11] Williams, "Nothing to Beat," 365.

[12] Clarke, Adams, and Barker, "Methods Matter," 360.

[13] Williams, "Nothing to Beat," 376, note 39.

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

In what follows, by "*Normal Clock*" I denote this current formulation.

Your true belief that it is 4:30 pm is sensitive to falsehood. Were it to be any time other than 4:30 pm when you observe the position of the hands of your clock, then you would not believe that it is 4:30 pm. Your true belief that it is 4:30 pm is also truth-adherent. Were you to observe the position of the hands of your clock at 4:30 pm while being slightly closer to it, then you would still believe that it is 4:30 pm. So far so good for Nozick's analysis, because surely you do know that it is 4:30 pm.

Then I *originally* described *Stopped Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Like *Normal Clock*, it has an analogue design so its hands are supposed to sweep its face continuously. However, it has no second hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because exactly twenty-four hours ago a stray fleck of dust chanced to enter the clock's mechanism, stopping it.[14]

In order to maximise parity with *Normal Clock*, here is my *current* formulation of *Stopped Clock*.

> You habitually nap between 4:00 pm and 5:00 pm. Your method of ascertaining the time you wake is to observe, between 4:00 pm and 5:00 pm, the position of the hands of your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However it has no second-hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because exactly twenty-four hours ago a stray fleck of dust chanced to enter the clock's mechanism, stopping it.

In what follows, by "*Stopped Clock*" I denote this current formulation.

Your belief that it is 4:30 pm is insensitive to falsehood. If it were not 4:30 pm but some other time, then by observing the position of the hands of your clock you would still believe—but then falsely—that it is 4:30 pm. This is more good news for Nozick's analysis, since surely you do not know that it is 4:30 pm. One

[14] Williams, "Nothing to Beat," 365.

very plausible explanation of your ignorance is that your belief is luckily true. You were lucky to look at the clock exactly twenty-four hours after it stopped working, at the only instant during the hour when you nap at which its hands could have pointed to the correct time.

I then *originally* described *Backward Clock* as follows.

> You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to look at your clock, one you know has always worked perfectly reliably. Unbeknownst to you, your clock is a special model designed by a cult that regards the hour starting from 4 pm today as cursed, and wants clocks not to run forwards during that hour. So your clock is designed to run perfectly reliably backwards during that hour. At 4 pm the hands of the clock jumped to 5 pm, and it has been running reliably backwards since then. This clock is analogue so its hands sweep its face continuously, but it has no second hand so you cannot tell that it is running backwards from a quick glance. Awaking, you look at the clock at exactly 4:30 pm and observe that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm.[15]

In their reply, Adams *et al* mounted objections to my argument that this example shows that Nozick's analysis fails. One of these was premised upon assumptions about the intentions of the cult in designing the clock.[16] In my rejoinder I pointed out that as a counterexample to Nozick's analysis, the intentions of its designers is an inessential feature.[17] I observed that;

> … we could dispense with the cult entirely and stipulate that a bug in the programming of the microchip circuit of your clock causes it run perfectly reliably backwards from 5:00 pm to 4:00 pm during a particular hour.[18]

In order to try to make this perfectly clear, and at the same time maximise parity with *Normal Clock* and with *Stopped Clock*, here is my *current* formulation of *Backward Clock*.

> You habitually nap between 4:00 pm and 5:00 pm. Your method of ascertaining the time you wake is to observe, between 4:00 pm and 5:00 pm, the position of the hands of your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However it has no second-hand. Awaking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because unbeknownst to you, the clock has continued to work perfectly reliably until 4.00 pm, when a bug in the programming of its microchip circuit caused its

[15] Williams, "Nothing to Beat," 366-367.
[16] Adams, Barker, and Clarke, "Beat the Clock," 357.
[17] Williams, "Nothing to Beat," Section 3 "The Irrelevant Intentions of the Cult", 370-372.
[18] Williams, "Nothing to Beat," 372.

> hands to jump to 5:00 pm and then run perfectly reliably backwards from 5:00 pm to 4:00 pm.

In what follows, by "*Backward Clock*" I denote this current formulation.

Your belief that it is 4:30 pm is luckily true, for you were lucky to observe the position of the hands of your clock at exactly 4:30 pm, at the only instant during the hour when you nap at which its hands could have pointed to the correct time. Thus you do not know that it is 4:30 pm any more than you do in *Stopped Clock*. Your belief also satisfies (4), in other words, is truth-adherent. If you were to observe the position of the hands of your clock at 4:30 pm while being slightly closer to it, then you would still believe that it is 4:30 pm.

But your belief that it is 4:30 pm is also sensitive to falsehood. If it were not 4:30 pm but some other time, then by observing the position of the hands of your clock you would not believe that it is 4:30 pm. Instead you would form *some other false belief* about what time it is. For example, if you were to observe the position of the hands of your clock at 4:31 pm, then you would not form the false belief that it is 4:30 pm. Instead you would form the false belief that it is 4:29 pm. Thus Nozick's analysis, as formulated above, is too weak, predicting knowledge where there is ignorance.

2. No Failure of Adherence

I will now examine Adams *et al*'s current objections, although not in any order they give them. This will prove to be a lengthy business. They now concede that your belief that it is 4:30 pm in *Backward Clock* is not sensitive to falsehood. Of your method of ascertaining the time that you wake, namely observing between 4:00 pm and 5:00 pm, the position of the hands of your clock (or as they more inaccurately put it "reading what the clock displays"[19]), they say the following.

> … the method employed will generate false beliefs and so your *beliefs* will be *insensitive* to the truth-value of p for all values of p *other than 4:30*. (my italics)[20]

Instead they attempt a new objection to my argument that *Backward Clock* counterexamples Nozick's analysis. In their reply they conceded that "Nozick's adherence condition … is not relevant to the example."[21] Now they repudiate their concession and argue that your belief in *Backward Clock* that it is 4:30 pm is not truth-adherent. They say that

[19] Clarke, Adams, and Barker, "Methods Matter," 102.

[20] Clarke, Adams, and Barker, "Methods Matter," 102.

[21] Adams, Barker, and Clarke, "Beat the Clock," 354, note 7.

> This is because satisfying the adherence condition requires something much stronger than mere true belief, it requires that if it were 4:30 (in other circumstances) then one would believe that it was 4:30 p.m.[22]

So far so good. We would all agree. But now they immediately make the startling claim, "But this is exactly what is not the case with the Backward Clock."[23]

We would do well to remember how I described *Backward Clock*. This was—in parity with the other two clocks—that you counterfactually observe the position of its hands while being *slightly closer* to it. These other circumstances or worlds are pretty close to the actual world as we have imagined it. The logical and physical changes we would have to make to them seem minimal. These count as 'slightly changed' circumstances. In these you still believe that it is 4:30 pm. They continue as follows.

> As we pointed out: "His belief is that it is 4:30, and it happens to be 4:30. But it is not the case that he believes it is 4:30 because it is 4:30-his believing it to be 4:30 is not explained by the fact that it is 4:30."[24]

They did indeed claim this.[25] What they now ignore is the fact that in my rejoinder I showed that this claim is false. I pointed out that;

> Your observation of the position of its hands, itself determined by their actual position, together with your understanding of how such positions represent time and your knowledge that your clock has always worked perfectly reliably, is what makes you believe that it is 4:30 pm.[26]

To elaborate on this point, in slightly changed circumstances in which you observe the position of the hands of your clock while being slightly closer to it than you actually are, the fact that it is 4:30 pm, together with actual temporal processes still in operation and the mechanism of your clock, causes its hands to point to 4:30 pm. In turn, its hands pointing to 4:30 pm, together with your reliable visual perception of their position and your knowledge of how such positions represent the time, causes you to believe, at 4:30 pm, that it is 4:30 pm. The formation at 4:30 pm of your belief that it is 4:30 pm is indeed to be explained, albeit partly, by the fact that it is 4:30 pm. This is all equally true of *Normal Clock*.

---

[22] Clarke, Adams, and Barker, "Methods Matter," 102.
[23] Clarke, Adams, and Barker, "Methods Matter," 102.
[24] Clarke, Adams, and Barker, "Methods Matter," 102.
[25] Adams, Barker, and Clarke, "Beat the Clock," 359.
[26] Williams, "Nothing to Beat," 373.

Alvin Goldman would agree. For him, *S* knows that *p* just in case the fact that *p* is causally connected in an "appropriate" way with *S*'s believing that *p*. The simplest appropriate causal connection is that the fact that *p* (that it is 4:30 pm) initiates a chain of causes that terminate in *S*'s believing that *p*.[27]

Besides, I fail to see what explanation has to do with the satisfaction of Nozick's adherence condition. To decide whether the adherence condition is satisfied we are to imagine slightly changed circumstances in which the content of *S*'s belief remains true and then decide whether *S* would still have that belief in these close-to-actual circumstances. No appeal to explanation is required.

Adams *et al* now continue as follows.

> The signal is too equivocal to be reliable in other circumstances since the clock might have been made not to read 4:30 even if it was 4:30 and so one would not believe that P though P is true.[28]

Again they talk of a signal as being 'equivocal', as they did in their reply.[29] Let us postpone examination of what this might mean until Section 5. A new idea has now appeared in their latest defence of Nozick, namely reliability. I will show in Section 6 that a reliabilist treatment of the truth-tracking analysis does them more harm than good. This leaves us with their claim that "the clock might have been *made* not to read 4:30 even if it was 4:30" (my italics). They now continue as follows.

> Suppose, for instance, the *clock shuts off* at 4:30 for one minute but otherwise reliably runs backwards from 5 until 4. You wake up, look up, see no time *on display*, and suspend judgement on the time for that minute. In such a possible world, P is true but you don't believe that P and so condition 4 is not satisfied. We deny, therefore, that condition four is satisfied concerning the Backward Clock Case (my italics).[30]

But these do not count as 'slight changes' to the circumstances that we were supposed to imagine as actual. The *actual* circumstances that we were supposed to imagine are those in which you *habitually* nap between 4:00 pm and 5:00 pm and observe the position of the hands of your *analogue* clock in order to ascertain the time you wake. So the circumstances that we are supposed to imagine as actual are those in which you habitually observe the positions of the hands of an *analogue* clock. You actually wake at 4:30 pm and observe the position of the hands. Adams

[27] Alvin I. Goldman, "A Causal Theory of Knowing," *Journal of Philosophy* 64, 12 (1967), 369.

[28] Clarke, Adams, and Barker, "Methods Matter," 102.

[29] For example, "…the clock's display is equivocal even if the clock wouldn't display '4:30' unless the time were 4:30" (Adams, Barker, and Clarke, "Beat the Clock," 358-359).

[30] Clarke, Adams, and Barker, "Methods Matter," 102.

et al now invite us to suppose that counterfactually, you wake at 4:30 pm and observe a *digital* clock that displays nothing at all. This is a pretty far out change of circumstances! You would be very surprised to discover that the analogue clock that you habitually use has suddenly changed to a digital clock. You would ask yourself what had happened to the hands. You might be well advised under such circumstances to down a stiff whiskey and then resume your nap—or if that does not help, to call a 24-hour horologist or even a psychiatrist. Who knows what you would believe or not believe?

Moreover, the actual circumstances we were supposed to imagine are those in which a bug in the programming of the microchip circuit of your clock caused its hands to jump to 5:00 pm and then run perfectly reliably backwards from 5:00 pm to 4:00 pm. Adams *et al* invite us to suppose that these actual circumstances are changed to those in which your clock was *made* such that at 4:00 pm it displays 5:00 pm and then runs perfectly reliably backwards from 5:00 pm until 4.30 pm when it stops until 4:31 pm, when it recommences running perfectly reliably backwards to 4:00 pm. These do not count as 'slight changes' to the circumstances that we were supposed to imagine as actual either. In the circumstances that we were supposed to imagine as actual, the clock is not *made* or *designed* to run backwards, but runs backwards due to a *bug* in the programming of its microchip circuit. To make this point more salient, I am free to stipulate that in all three clocks, the mechanism was designed to operate as *Normal Clock*, but that in *Backward Clock*, exactly twenty-four hours before you wake at 4:30 pm, a stray fleck of dust chanced to enter the clock's mechanism initiating corrosion to its microchip circuit that created a bug in its programming that caused its hands to jump to 5:00 pm and then run perfectly reliably backwards from 5:00 pm to 4:00 pm. On this stipulation, the mechanisms of both *Backward Clock* and *Stopped Clock* actually operate the way that they do, *not by design but by luck*. Talk of design or intentions of designers is simply irrelevant.

I asked you to suppose that instead of observing the position of the hands of your clock while being close to it with its *Backward Clock* mechanism unchanged, you are counterfactually slightly closer to it. In stark contrast, Adams *et al* invite you to suppose that that instead of observing the position of the hands of your clock while being close to it with its *Backward Clock* mechanism unchanged, its mechanism is counterfactually changed. The mechanism is to be changed to a combination of the mechanisms of *Backward Clock* and *Stopped Clock* with *Stopped Clock* sandwiched for a minute at 4.30 pm inside *Backward Clock*. Let us call a clock with this mechanism *Stopped Clock Sandwiched in Backward Clock*. But this is not changing the mechanism of *Backward Clock* but changing

Backward Clock to a different clock, namely *Stopped Clock Sandwiched in Backward Clock*. After all, a clock is essentially its mechanism. This is why they bear names such as "*Normal Clock*," "*Stopped Clock*" and "*Backward Clock*" rather than say, "*Red Clock*," "*Blue Clock*" or "*Green Clock.*" Adams *et al* are inviting us to change the *actual* case we were supposed to imagine. But to test adherence we need only make *counterfactual* changes to the case we were supposed to imagine as actual. Adams *et al* do not have the luxury to change my examples to suit them.

Another way to look at this is that Adams *et al* invite us to start with the actual circumstances of *Backward Clock*, but in deciding whether the adherence condition is satisfied, change *Backward Clock* to *Stopped Clock Sandwiched in Backward Clock* while still calling it *"Backward Clock."* While calling them by their original names, they shuffle different clocks under those names. This is reminiscent of the trick with three upside-down cups. The trickster shows these as empty and then places one over a coin and shuffles the cups, invariably leading his dupe to guess incorrectly which cup covers the coin.

Besides, if Adams *et al* are allowed to change the mechanism in testing adherence in *Backward Clock*, then I should be allowed to do the same in *Normal Clock*. I originally invited you to suppose that instead of observing at 4:30 pm the position of the hands of your clock while being close to it with its forward-running mechanism unchanged, you are counterfactually slightly closer to it. But now it seems, I am allowed to suppose that instead of observing at 4:30 pm the position of the hands of your clock while being close to it with its forward-running mechanism unchanged, you observe at 4:30 pm the position of the hands of your clock with its mechanism counterfactually changed. In these 'slightly changed' circumstances it might have a mechanism that stopped at 4:15 pm or a mechanism that at 4:00 pm made the hands of your clock jump to 5:00 pm and then run backwards more slowly so that its hands never point to the correct time during the hour that you nap. In these circumstances the hands do not point to 4:30 pm when at 4:30 pm you observe the position of its hands. So you do not believe that it is 4:30 pm. Adherence fails. This predicts ignorance where there is knowledge. Two can play at the three-cup trick! As Sinhababu and I said in our original paper:

> Of course the closeness of possible worlds to actuality is vague, but close possible worlds cannot include those in which the *mechanism* of the clock differs from its *actual* mechanism. This is because the truth-adherence of your belief that it is 4:30 pm in *Normal Clock* resides in the fact that you would still have that belief in slightly changed circumstances in which the mechanism of the clock continues to work perfectly reliably. Likewise, the worlds close to the actual

> circumstances of *Stopped Clock* surely include those in which the mechanism of the clock is stopped.[31]

We are now in a position to recap. Adams *et al* admit that you do not know that it is 4:30 pm in *Backward Clock*.[32] They admit that your belief that it is 4:30 pm is true.[33] As just shown, they now concede that it is sensitive to falsehood. As just shown, their objection that it is not truth-adherent fails. It is indeed truth-adherent. This is enough to show that Nozick's analysis, as formulated above, is too weak. Nonetheless, let us examine their remaining objections.

3. No Mistake About Methods

Next is a set of objections that I have made various mistakes about methods. Before examining these, some clarification is helpful. In our original paper Sinhababu and I took sensitivity (and indeed adherence and truth-tracking) to be a condition on *S*'s *belief* that *p*, not on her *method M* of arriving at that belief. In so doing we followed Nozick and his commentators.[34] In their reply, Adams et al talked of *sensitive methods* and *truth-tracking methods*, insisting that your method of ascertaining the time you wake is not sensitive or truth-tracking. But a careful reading of *Philosophical Explanations* reveals that every time that Nozick talks of sensitivity, he talks of *beliefs* or *belief states* as being sensitive.[35] He does

---

[31] Williams and Sinhababu, "The Backward Clock," 49.

[32] "… we think that the method, i.e., 'looking at the clock and determining what it says,' is too equivocal to yield knowledge" (Clarke, Adams, and Barker, "Methods Matter," 101).

[33] "…the person awaking from the nap happens to acquire a true belief that it is 4:30" (Adams, Barker, and Clarke, Beat the Clock," 359).

[34] Williams, "Nothing to Beat," 365, note 5.

[35] As far as I can tell, here is an *exhaustive* list of the passages in which Nozick talks of sensitivity in *Philosophical Explanations*. Italics are my own. "The person in the tank does not know he is there, because his *belief* is not sensitive to the truth." (175); "The subjunctive condition (3) … tells us only half the story about how his *belief* is sensitive to the truth-value of *p*. It tells us how his *belief state* is sensitive to *p*'s falsity, but not how it is sensitive to *p*'s truth; it tells us what his *belief state* would be if *p* were false, but not what it would be if it were true." (176); "To be sure, conditions l and 2 tell us that *p* is true and he does believe it, but it does not follow that his *believing p* is sensitive to *p*'s being true." (176); "His *belief* is not sensitively tuned to the truth, he doesn't satisfy the condition that if it were true he would believe it." (177); "We do not mean such a person to easily satisfy 4, and in any case we want his *belief-state*, sensitive to the truth of *p,* to focus upon *p*."(178); "Once we have the notion of a *belief* varying with or being sensitive to the truth-value of what is believed, we see there are differing degrees of such sensitivity or covariation." (283); "It makes their beliefs (sometimes) vary somehow with the truth of what is believed; it makes their *beliefs* somehow sensitive to the facts." (285).

not talk of *methods* as the sort of things that can be sensitive. Every time that he talks of truth-tracking, he talks of *beliefs* that are truth-tracking.[36] He does not talk of *methods* as the sort of things that can be truth-tracking.

In my rejoinder, I pointed out that Adams et al owe us an analysis of 'sensitive methods.' I offered them a trilemma; a sensitive method is one that sometimes, mostly or always produces sensitive beliefs. I then demonstrated that accepting any option leads to disaster.[37] In their latest defence of the tracking theories they fail to mention this objection. Instead, they drop talk of 'sensitive

[36] As far as I can tell, here is an *exhaustive* list of the passages in which Nozick talks of tracking in *Philosophical Explanations*. Italics are my own. "A closest relative version of this last theory, as we shall see, holds that a person knows, via method M, that *p,* if his *belief*, via M, that *p* tracks the fact that *p,* and if there is no other method M′ outweighing M but not tracking the truth, via which also he believes that *p*." (49); "*Belief* that *p* tracks truth that *p*" (52); "What we have when our beliefs vary subjunctively with the truth is knowledge, but if there were beings whose beliefs varied more closely and extensively with the truth of what they believed, beings whose *beliefs* did more than track, in that case our beliefs which are knowledge would not be knowledge." (54). I admit that the passage at p.49 suggests that there could be a method M′ outweighing M but not tracking the truth. This in turn suggests that M′ is the sort of thing that can be truth-tracking. On the other hand, to achieve coherence with the body of his other remarks above, the last conjunct of this outlier could be charitably read as "… and if there is no other method M′ outweighing M *that does not produce beliefs that track the truth*, via which also he believes that *p*." Besides, if talk of truth-tracking methods is allowed to peep into the picture in the special case in which *outweighing* methods are involved, it does not follow that they have any place in the more straightforward cases where *single* methods are involved, which as Adams *et al* admit, constitute the arena of our controversy (Clarke, Adams, and Barker, "Methods Matter," 100).

[37] Williams, "Nothing to Beat," Section 6 "What Are Truth-tracking Methods?" 375-378. Formulating a sensitive method as one that *sometimes* produces sensitive beliefs does not help them because the method you use in *Backward Clock*, namely observing the positions of its hands during the hour you nap, produces a sensitive belief when you use it at 4:30 pm. Formulating it as one that *mostly* produces sensitive beliefs succumbs to *Recently Stopped Clock* as follows. You habitually nap between 4 pm and 5 pm. Your method of ascertaining the time you wake is to observe the position of the hands of your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. However, it has no second hand. Awaking at 4:55 pm, you see that its hands point to 4:55 pm. Accordingly, you form the belief that it is 4:55 pm. And it is indeed 4:55 pm because the clock has continued to work perfectly reliably until 4:50 pm, when a bug in the programming of its microchip circuit caused its hands to jump to 4:55 pm and then stop. *Most* of the beliefs that you might form by observing the positions of the hands of your clock during the hour that you nap are sensitive. But you do not know that it is 4:55 pm. Now suppose instead, that waking at 4:30 pm, you see that its hands point to 4:30 pm. Accordingly, you form the true belief that it is 4:30 pm. Surely you know that it is 4:30 pm, but not *all* of the beliefs that you might form by observing the positions of the hands of your clock during the hour that you nap are sensitive.

methods,' allowing it to slip in just once at the end of their paper.[38] Instead they start to talk of methods as being 'reliable.'[39]

Of course, the fact that for Nozick, sensitivity is a property of belief, and not methods is perfectly compatible with the fact that for him, beliefs are sensitive because they are formed *via* a method. That *X* has property *F* because *X* is formed *via M*, does not entail that *M* is *F*. Your alert appearance, formed because you methodically sleep early, does not mean that your sleeping early enjoys alert appearance.

Two other clarifications are in order. First, Nozick's analysis, as formulated above, takes as its *definiendum*, "*S* knows that *p*, using method *M* of arriving at a belief whether *p*." Adams *et al* now pose the following rhetorical question.

> … how could condition (4) be satisfied by anyone, if methods aren't the means? Truths don't just pop into heads. It often takes hard work (science, detectives) to discover them. We take it as an obvious fact about tracking theories (Nozick's or Dretske's) that beliefs only track in virtue of reasons or methods. Otherwise, such theories would make no sense.[40]

I am not so sure that truths don't sometimes just pop into our heads. Why can't you hit on a truth as the result of accidental discovery, divine enlightenment, insight or the agency of Sherrilyn Roush's *Fairy Godmother* who ensures that all your beliefs are true?[41] Fortunately I may enjoy neutrality on the question of whether there can be knowledge involving beliefs that are not formed using *any* methods, pre-reflective, unconscious, applied un-methodically, incapable of articulation, formed from the inside or whatever. But if there can be such knowledge, then that means, not that Nozick's analysis is senseless, but merely that it is incomplete.

Second, it is transparently obvious that the same method *M* reappears in (3) and (4) of the *definiens*. So for Nozick, *S*'s belief is sensitive to falsehood just in case by using the *same* method of belief-formation that she *actually* uses, she would not have that belief if it were false. And for him, her belief is truth-adherent just in case by using the *same* method of belief-formation that she

---

[38] "Those *methods* M or reasons R must be *sensitive* for both Dretske and Nozick and additionally adherent for Nozick in order to track the truth" (Clarke, Adams, and Barker, "Methods Matter," 112, my italics).

[39] For example, "That method must be absolutely reliable with respect to a variety of input beliefs in near possible worlds for *S* to know that *p*." (Clarke, Adams, and Barker, "Methods Matter," 102).

[40] Clarke, Adams, and Barker, "Methods Matter," 106.

[41] Sherrilyn Roush, *Tracking Truth* (Oxford: Oxford University Press, 2005), 122-123.

actually uses, she would still have that belief if it were true under slightly changed circumstances. This means that in deciding whether her belief—one she actually forms by a method—tracks the truth, we must hold that method *fixed* across close possible worlds. That is agreed on all sides. Nonetheless, for Nozick, it is still the belief that is sensitive, truth-adherent or truth-tracking, not the method.

With these clarifications in place, we may now turn to Adams *et al*'s objections that I am mistaken about methods. They say the following.

> Our mistake, on William's view, is mistaking Nozick's sensitivity condition as a constraint on **METHOD** rather than **BELIEF**. But it is Williams who misunderstands Nozick's theory, not us. This is because it is exactly the method that ensures that the correct connection between belief and fact obtains when we know some factual belief, that the belief is both sensitive and adherent.[42]

I have already shown why this gets things wrong. Even if it is true that a belief is sensitive only in virtue of being produced by a method, it does not follow that it is the method that is sensitive. Nozick never talks of sensitive methods. He only talks of sensitive beliefs. If Adams *et al* wish to introduce the notion of sensitive methods then they part company with Nozick. And at that point of departure they owe us an analysis of what a sensitive method is supposed to be, one that avoids my trilemma.

Later in their current defence, they say the following.

> By proudly announcing his intention to focus only upon beliefs, Williams guarantees non-success at responding to our reply to his Backward Clock Case. In this context, Williams' claim that we are defending a different theory than Nozick's because we talk about **METHOD** fails.[43]

I announced no such thing, proudly or otherwise. For Nozick, *S*'s belief is sensitive to falsehood just in case by using the same method of belief-formation that she actually uses, then she would not have that belief if it were false. This does not put the focus only on beliefs. It mentions methods as well. Nonetheless, for Nozick, it is the *belief* that is sensitive, not the *method*. Nor did I claim that Adams *et al* are defending a different theory than Nozick's because they talked about *methods*. I claimed that they are defending a different theory than Nozick's because they talked about *sensitive methods*.[44] Why can't they tell us what these are supposed to be?

---

[42] Clarke, Adams, and Barker, "Methods Matter," 102.

[43] Clarke, Adams, and Barker, "Methods Matter," 106.

[44] "This might make us suspect that Adams, Barker and Clarke are defending a different tracking theory from Nozick's. This is further confirmed by the fact that they argue that you do not have a truth-tracking method of forming the belief that it is 4:30 pm. But Nozick's analysis is not

Their final objection that I have made a mistake about methods proceeds as follows.

> What is crucial then, for Williams, is that the mechanism of the clock is held fixed across close possible worlds when considering subjunctive conditionals of the sort that Nozick imposes on knowledge. Unfortunately for Williams and Sinhababu, what needs to be held fixed across possible worlds is not **mechanisms** but the **method M** for Nozick, or, for Dretske, the **circumstances C** relative to the reasons or evidence R.[45]

Adams *et al* appear to commit the fallacy of false dichotomy. They seem to reason that either methods or mechanisms are to be held fixed, and since methods are to be held fixed, mechanisms are not to be held fixed. That is like arguing that Trump is either in Saudi Arabia or in a meeting with Russian intelligence, and since he is in Saudi Arabia, he is not in a meeting with Russian intelligence. He could easily be in a meeting with Russian intelligence in Saudi Arabia. As I have just shown, in deciding whether *S*'s belief tracks the truth, one she actually forms by a method, we must hold that method *fixed* across close possible worlds. That is not in dispute. It never was. And as I have already shown in Section 2, in deciding whether your belief that it is 4:30 pm tracks the truth in *Normal Clock*, *Stopped Clock* or *Backward Clock*, we must *also* hold the actual mechanism of the clock fixed across close possible worlds.

4. The Appeal to Charity and Extra-Sensitivity

One of Adams *et al*'s more interesting objections to my rejoinder is an appeal to charity. This goes as follows.

> The Backward Clock Case would have led to a quick death for Nozick's theory within minutes of his thinking of the theory if he had understood his own theory as Williams does. Why? Because Nozick himself would have understood that the theory, understood in that way, was bankrupt! But, says Williams, mightn't Nozick not have noticed that the theory has this very odd consequence that even false 'not *p*' beliefs can serve to confirm condition three of the theory for a particular *p*? The correct answer to this objection is: No, only an extremely uncharitable reading of the theory could possibly interpret Nozick as intending, or leaving open, or suggesting, or not noticing, this interpretation of the theory. The principle of charity counsels us to avoid implausible and unlikely

elucidated in terms of a truth-tracking method, but in terms of a truth-tracking belief" (Williams, "Nothing to Beat," 368).

[45] Clarke, Adams, and Barker, "Methods Matter," 108.

> interpretations of the words of an author. If ever there was an implausible and unlikely reading of a theory, Williams' reading of Nozick is it.[46]

Before examining this objection, we first need to clarify the rather obscure locution "'not *p*' beliefs." We should not read this as referring to a belief that not *p*, which in *Backward Clock* would be your belief that it is not 4:30 pm. Rather, we should read it as referring to beliefs *other than* the belief that *p*, which would be your beliefs other than that it is 4:30 pm, such as your belief that it is 4:29 pm. This is corroborated by earlier passage that runs as follows.

> After all, the very idea of 'not p' includes, among other things, the entire universe of visual perception beliefs other than *p*![47]

Now the "very odd consequence" becomes as follows. If it were not 4:30 pm when you observe the position of the hands of *Backward Clock* but say 4:31 pm, then your false belief that it is 4:29 pm, formed at 4:31 pm, "serves to confirm" the sensitivity of your belief that it is 4:30 pm formed at 4:30 pm. That however, is not an entirely accurate thing to say of *Backward Clock*. Rather, what makes your actual belief that it is 4:30 pm sensitive is that if it were not 4:30 pm when you observe the position of the hands, but say 4:31 pm, then you would not believe that it is 4:30 pm. It is also true that you would form some other false belief, in this case, the false belief that it is 4:29 pm. But that is not the source of sensitivity. This is how I properly understand matters.

Now we can address the appeal to charity. I agree that if Nozick had understood matters in the same way when presented with *Backward Clock*, then he would have seen that his theory is bankrupt. But we should not forget that Nozick was *not* presented with it. Even he had been, it might have taken him a while to recognize that your belief that it is 4:30 pm is sensitive in exactly his own sense of 'sensitive.' After all, Adams *et al* have taken a very long time to reluctantly concede this point. There are plenty of theories that were seen to be untenable only once the right objection was made and reflected on.[48] Who is to say that *Backward Clock* is not such an objection? It took me quite a while myself to see that it refutes Nozick! Charity goes both ways. In fact we should be even more charitable to our opponents than those we support.

---

[46] Clarke, Adams, and Barker, "Methods Matter," 104.

[47] Clarke, Adams, and Barker, "Methods Matter," 104.

[48] One apposite case is Gettier's counterexamples to the analysis of knowledge as justified true belief. Others include Russell's paradox and the inconsistency of Frege's Axiom V, Gödel's incompleteness theorem and Hilbert's program or Lewis' triviality proof and Stalnaker's thesis.

I freely admit that I may have misunderstood Adams *et al*. Maybe by "reading of the theory" they do not refer to sensitivity as formulated by Nozick in condition (3). Before their appeal to charity they say the following.

> Hence, the idea that one must track the truth of 'not p' veridically demonstrates that the method must be absolutely reliable for beliefs of that type …When Nozick asserts condition three his point is that, given your method, you would track the truth of 'not p' veridically, whether the resulting belief be *q*, *r* , or *s* …. One must not believe that *p* and veridically track the truth such that you don't believe that, say, *q* falsely.[49]

Given our clarification of "'not *p*' beliefs," this appears to attribute to Nozick the view that to know that *p* by a method of forming the belief that *p*, you would not believe that *p* if it were false that *p*, *nor would you form any other false belief of the same type*. Adams *et al* could now accuse me of uncharitably not reading Nozick this way. The idea behind the proposed reading is that not only must your belief be sensitive to its falsehood, but must also be resistant to producing other false beliefs of the same type in the neighbourhood as well. To coin a term, it must be *extra-sensitive* to falsehood, as follows.

> *S*'s belief that *p* is extra-sensitive just in case were it false that *p*, then *S* would not believe that *p* and would not acquire a false belief of the same type other than the belief that *p*.

It follows that extra-sensitive beliefs are sensitive but not necessarily conversely.

I see no textual basis for attributing this to Nozick. But if it is his view, so much the worse for him. Of course, requiring extra-sensitivity for knowledge blocks *Backward Clock*. It also explains why you do not know that it is 4:30 pm in *Stopped Clock*, since your belief that it is 4:30 pm is not sensitive. Extra-sensitivity also accommodates *Normal Clock*. Were it any time other than 4:30 pm when you observe the position of its hands, then you would not believe that it is 4:30 pm. Nor would you form another false belief about the time. Alas, substituting extra-sensitivity for sensitivity proves too much, as shown by *Fake Dog Occluded by Sheep*.

> Directly in front of you where you are looking is a sheep and directly behind it, occluded from your vision by the sheep, is a fake dog that looks just like a dog. Using your reliable vision and memory in ordinary circumstances, you believe that what is directly in front of you is a sheep.

[49] Clarke, Adams, and Barker, "Methods Matter," 104.

Your belief that what is directly in front of you is a sheep, is sensitive but not extra-sensitive. Were there to be no sheep directly in front of you, then you would not believe that what is directly in front of you is a sheep. Instead you would acquire the *false* belief that what is directly in front of you is a dog. That false belief is of the same type as your belief that what is directly in front of you is a sheep, being visual-cum-memorial as well as about the species of the animal directly in front of you. But surely you know that what is directly in front of you is a sheep. So now the analysis is too strong, predicting ignorance where there is knowledge.

5. 'Equivocation' and the Boy Who Cried "Wolf!"

In their reply, Adams *et al* made the following objection.

> For any time other than exactly 4:30, the subject's belief during that hour-long period will be false. Why? Because the clock *lies* for all but one moment during that hour-long period. And worst of all, there is nothing in the signal sent by the clock to differentiate when it is telling the false time from when it is telling a true time.
>
> This should remind one of the "little boy who cried 'wolf.'" The boy cries 'wolf' over and over when there is no wolf. Then on the one occasion when there is a wolf and he cries 'wolf,' his cry has become *to equivocal*, no one can tell from his cry that a wolf is actually there on that one occasion. His cry of 'wolf' still means wolf, but it does not carry the information that there is a wolf. Similarly, the clock's face emits false testimony for 59 minutes during that hour from 4:00 to 5:00 (my italics).[50]

Now they lodge the following complaint.

> The 'Boy who cried Wolf' Case, which we developed at length in our reply, was expressly devised to make the point that equivocal signals will not generate knowledge. Conveniently, Williams ignores this argument from our response.[51]

I confess that I did indeed ignore this response. Dealing with it was in a sense inconvenient, because I could not see its relevance to the status of *Backward Clock* as a counterexample to the tracking analysis. Very well, here is why it is irrelevant.

First, if you were to observe the position of the hands of your clock at any time other than 4:30 pm, say 4:31 pm, it would not tell you a lie. The clock has no intention to deceive you into believing falsely that it is 4:29 pm. Clocks are not the

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

[51] Clarke, Adams, and Barker, "Methods Matter," 104-105.

sort of things that have intentions. Nor would its designers have deceitful intentions, since its mechanism actually operates the way that it does, not by design but by luck. Moreover, now Adams *et al* claim that *Backward Clock* and *Stopped Clock;*

> … don't really **SAY** anything about the time, even though they continue to **DISPLAY** the time and appear to **SAY** something about it.[52]

But if your clock doesn't say anything, then it doesn't tell you anything and so does not tell you a lie. Adams *et al* can't have their cake and eat it.

Second, it is true that there is nothing *in the signal*, in other words, the position of the hands of *Backward Clock*, to differentiate when it is telling the correct time from when it is not. Although you are justified in believing that the hands of *Backward Clock* point to the correct time (because you know that it has always worked perfectly reliably), you cannot be absolutely sure, *just by observing the position of the hands*, when these point to the correct time. To know that, you would have to use induction from its unfailing past accuracy or an independent check of its accuracy, such as another clock that you know is accurate. But this is equally true of *Normal Clock*. You cannot be absolutely sure, *just by observing the position of the hands*, when these point to the correct time. So if this is an impediment to knowledge in *Backward Clock*, then it is equally an impediment to knowledge in *Normal Clock*. It isn't.

Third, Adams *et al* appear to use the term 'equivocal' in a deviant way. The normal understanding of equivocation is that it takes place just in case the *same* bearer of meaning, such as the *same* word, phrase, or signal occurs at different times with different meanings. If the boy were to call "Bank!" at one time and mean "There is the side of a river!" and call "Bank!" at another time and mean "There is a financial institution!," then that would be equivocation. But in *Backward Clock*, *different* positions of its hands, in other words *different* signals, represent different times, or roughly, have different meanings. Moreover, as Adams *et al* themselves admit, the different times at which the boy cries "Wolf!" are those at which his cry means the *same* thing, roughly "There's a wolf!" Since Adams *et al* rest their objection on an appeal to 'equivocation,' they owe us an explanation of what this is supposed to be. Otherwise we are in no position to properly assess the objection.

Fourth, their analogy between *Backward Clock* and the story of the boy who cried "Wolf," is weak. In the story, the boy cries "Wolf!" a series of times. Each time you wait a while and confirm that there is no wolf. So you know, after

[52] Clarke, Adams, and Barker, "Methods Matter," 110.

each time, that what the boy cries is *false*. Then the boy again cries "Wolf!" and there is a wolf although you have yet to confirm or disconfirm this. At that juncture you are justified in believing that what boy cries is *false*, although it is true. You have sensibly used induction using the evidence available to you. This is a significant disanalogy with *Backward Clock*. Because you know that it has always worked perfectly reliably, you know after all past times at which you observed the position of its hands that these pointed to the *correct* time. Then you wake during the particular hour that its hands run backwards and observe its hands pointing to 4:30 pm. At that juncture, you have no idea that it is running backwards. You are justified in believing that its hands point to the *correct* time,[53] and they do point to the correct time. You have sensibly used induction using the evidence available to you. You should not trust the boy, but you should trust your clock!

Finally, even if it is true that "equivocal signals will not generate knowledge," or in other words that your method of ascertaining the time you wake in *Backward Clock* is 'equivocal' in such a way that prevents you from knowing at 4:30 pm that it is 4:30 pm, then that is grist to my mill. The fact remains that your belief that it is 4:30 pm is both sensitive to falsehood and truth-adherent. In their reply, Adams *et al* tried to evade this result by claiming that your method of ascertaining the time you wake, is 'equivocal' in such a way that prevents your *method* from being truth-tracking, and knowledge requires truth-tracking *methods*.[54] As we saw above in Section 3, here they part company with Nozick, and here they owe us an analysis of knowledge in terms of truth-tracking *methods*, first elucidating these while avoiding my trilemma. Now they have dropped talk of *truth-tracking* methods in favour of talk of *reliable* methods. So

[53] It might objected that in *Backward Clock* you are not justified in forming any belief about what time it is by observing the position of its hands during its backward-running hour, because to be so justified you would have to check that its hands are still moving forwards. As Adrian Heathcote formulates this claim, "Knowing the time by looking at a clock is a matter of confirming that it is still running" ("Gettier and the Stopped Clock," *Analysis* 62, 2 (2012): 309-314, 312). I think that this puts the bar too high in most cases, although it is plausible that to be justified in believing that the time is what the hands of the clock point to, you need to be justified in believing that the clock is still working reliably. In fact we only occasionally confirm that our clocks are still working reliably. These occasions sustain our justified confidence that they have continued to work reliably on the much more frequent occasions on which we only glance at them and hence normally tell the time, in other words, gain both knowledge and justified belief of what time it is.

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

perhaps they wish to argue that your method of ascertaining the time you wake is 'equivocal' in such a way that prevents your method from being *reliable*, and knowledge requires reliable methods. Let us now examine this line of thought.

6. The Appeal to Reliable Methods

The term 'reliable' does not appear in Nozick's analysis, as formulated above. In what sense does knowledge require reliable methods for Adams *et al*? An answer is found in their discussion of Nozick's *Grandmother*.[55] They say the following.

> … all visual perception beliefs, *all* beliefs of that type about her Grandson, must be reliably produced by that method for the Grandmother under those circumstances in order for her to know that *p*. (my italics)[56]

This commits them to the following condition on knowledge, which I dub '*ABC reliability*.'

> *S* knows that *p*, using method *M* of arriving at a belief whether *p* of type *T*, only if all beliefs of type *T* are reliably produced by *M*.

By 'under those circumstances' they envision slightly changed circumstances or close possible worlds. They also say the following.

> Tracking the truth presupposes the **reliability** of the **method** for producing truth. Reading his account in any other way is simply a misreading of Nozick.[57]

Presumably Adams *et al* do not propose to add *ABC reliability* as an extra condition in Nozick's analysis of knowledge, since they suppose it to be already entailed by conditions (3) and (4).

But there is no textual evidence that Nozick embraces *ABC reliability*. Indeed there is textual evidence that he does not. Long after Nozick has finished elucidating his truth-tracking analysis of knowledge,[58] he turns to a different

---

[55] A grandmother sees her grandson is well when he comes to visit but if he were too unwell to visit, then relatives would tell her that he is well to spare her upset. She arrives at the true belief that he is well *via* the method of looking at him, yet if he were unwell then she would still believe that he is well *via* the different method of testimony. So without mention of methods, (3) is false, but nonetheless she knows that he is well (*Philosophical Explanations*, 179). Hence the need in such 'multiple methods' cases to hold the method fixed from actuality across close possible worlds when testing for sensitivity or adherence, as reflected in Nozick's analysis as formulated above.

[56] Clarke, Adams, and Barker, "Methods Matter," 103.

[57] Clarke, Adams, and Barker, "Methods Matter," 104.

[58] Nozick, *Philosophical Explanations*, 172-185.

topic, namely evidence and justification.[59] Here he mentions reliability for the very first time, not as part of an analysis of *knowledge*, but as part of an analysis of *justified belief*. He holds that *S*'s belief that *p*, using method *M*, is justified if *M* is reliable, in other words, "is likely to produce mostly true beliefs."[60] Let us call this condition '*Nozick reliability*' as follows.

> *S*'s belief that *p*, using *M*, is justified if *M* is reliable, that is, is likely to produce mostly true beliefs.

For Nozick, truth-tracking is a property of belief, while reliability is a property of methods. In passing, he ponders "a stronger notion of reliability, one wherein the application of a method reliably yields knowledge (tracking) rather than simply truth."[61] He never pursues this notion. Indeed it is difficult to see how we are supposed to derive *ABC reliability* from *Nozick reliability.*

Nozick also tells us the following.

> When tracking holds, if it is true (false) you would (not) believe it-when reliability holds, if it is believed (by the method) then it (probably) would be true. It is important to keep these directions distinct.[62]

It is far from clear that either direction entails the other. In *Backward Clock*, your method of forming beliefs about the time is unreliable; since you can wake at any time during the hour that you nap, it is not likely to produce *mostly* true beliefs. Nonetheless by using that method at 4:30 pm you form a belief that tracks the truth.

Second, *ABC reliability* falls prey to much the same example that defeats an appeal to truth-tracking methods. This is a clock that combines *Normal Clock* with *Stopped Clock*, namely *Recently Stopped Clock*, as follows.

> You habitually nap between 4:00 pm and 5:00 pm. Your method of ascertaining the time you wake is to observe, between 4:00 pm and 5:00 pm, the position of the hands of your clock, one you know has always worked perfectly reliably. This clock is analogue so its hands sweep its face continuously. Awaking at 4:30 pm, you observe that its hands point to 4:30 pm. Accordingly, you form the belief that it is 4:30 pm. And it is indeed 4:30 pm because the clock has continued to work perfectly reliably until 4.50 pm, when a bug in the programming of its microchip circuit caused its hands to jump to 4:55 pm and then stop.

---

[59] Nozick, *Philosophical Explanations*, 264-268.

[60] Nozick, *Philosophical Explanations*, 264. Nozick adds "without holding that some particular degree of reliability is either sufficient or necessary for the beliefs being justified." (265).

[61] Nozick, *Philosophical Explanations*, 264, note *.

[62] Nozick, *Philosophical Explanations*, 266.

Not *all* of the beliefs that you might form by the method of observing the positions of the hands of your clock during the hour that you nap are reliable. These are those that you would form during the period from 4:50 pm to 5:00 pm when it functions as *Stopped Clock*. During that period your method is unlikely to produce mostly true beliefs and so does not reliably produce them. You use the very same method at 4:30 pm to arrive at the belief that it is 4:30 pm. As with all the other beliefs, you form it by observing the position of the hands of the clock, remembering how such positions represent time, recalling that the clock has always worked perfectly reliably, and inducing from all this what time it is. Thus *ABC reliability* predicts that you do not know that it is 4:30 pm. Surely you do, because that is what you know in *Normal Clock*—which from 4:00 pm to 4:49 pm is essentially the same as your clock. So now the analysis is too strong, predicting ignorance where there is knowledge.

Finally, Adams *et al* are now ensnared by the generality problem.[63] For example, they tell us the following.

> It is exactly because the Grandmother's method, i.e., visual perception, is absolutely reliable for close distances to her in near possible worlds, that she can confidently say that her Grandson is well.

But what is her method of forming the belief that her grandson is well? That of perception, visual perception, visual perception at close distances, visual perception at close distances using spectacles, visual perception of people at close distances using spectacles, visual perception of relatives at close distances using spectacles, keen observation of possible symptoms of illness in grandsons at close distances using spectacles, keen observation of possible symptoms of illness in grandsons at close distances using spectacles after a stiff whiskey and a nap, or some other type of method? The problem is that the reliability of the method may vary with how narrowly or broadly its type is described. How do we decide which repeatable type of method to select in a way that is not *ad hoc* in order to determine the reliability of the method that she actually uses?

7. Why Dretske's Early Analysis Remains Clocked Out

I *originally* formulated Dretske's early analysis of knowledge as follows.

> *S* knows that *p* just in case
>
> (1) S believes that *p* (without doubt, reservation or question) on the basis of *R*.

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

(2) *R* would not be the case unless *p* were the case.

(3) Either *S* knows that *R*, or *R* is some experiential state of *S*.[64]

This formulation should be made more precise. Dretske tells us the following.

> The circumstances which are assumed constant, which are tacitly held fixed, in conditionals such as (2), are those circumstances prevailing on the occasion in question (the occasion on and between which the particular states R and P obtain) which are logically and causally independent of the state of affairs expressed by P.[65]

To reflect this, here is my *current* formulation.

S knows that *p* in circumstances *C* just in case

(1) S believes that *p* (without doubt, reservation or question) on the basis of *R*.

(2) In *C, R* would not be the case unless *p* were the case.

(3) Either *S* knows that *R*, or *R* is some experiential state of *S*.

Dretske's *Thermometer* illustrates this nicely.

> In circumstances in which you have shaken your mercury thermometer down, and in which it does not stick, you place it in your child's mouth, extract it after several minutes and observe a reading of 98.6 F. You know that the thermometer reads 98.6 F. On this basis you believe without doubt, reservation or question, that the temperature is 98.6 F.[66]

This is supposed to explain why you know that the temperature is 98.6 F in circumstances in which you have shaken your mercury thermometer down and in which it does not stick. In these circumstances, the thermometer would not read 98.6 F unless the temperature was 98.6 F. In these circumstances, the thermometer is *reliable as an indicator of actual temperature*. The rise and fall of its column of mercury is deterministically and predictably correlated with the temperature it indicates. So it is also *reliable in the way its mechanism operates*. These circumstances are logically and causally independent from the temperature being 98.6 F. The fact that your thermometer is shaken-and-not-sticky neither entails nor causes the actual temperature to be 98.6 F. Conversely, the actual

[64] Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49, 1 (1971): 12–13. I gave this formulation in Williams, "Nothing to Beat," 374.

[65] Dretske, "Conclusive Reasons," 7-8.

[66] This is a slight embellishment, entirely in his spirit, of Dretske's example in "Conclusive Reasons," 2.

temperature being 98.6 F neither entails nor causes your thermometer to be shaken-and-not-sticky. So far so good for Dretske.

But now let us modify *Backward Clock* slightly. Suppose that as you observe the position of its hands, you believe that it is 4:30 pm without doubt, reservation or question, because you know that your clock has always worked perfectly reliably. You base that belief upon your conjunctive reason that the hands point to 4:30 pm and your clock has always worked perfectly reliably. But this conjunction would not be true unless it were 4:30 pm, because the hands would not point to 4:30 pm unless it was 4:30 pm. This is because the circumstances in which you find yourself include those in which the clock runs perfectly reliably backwards from 5:00 pm to 4:00 pm. Finally, we may stipulate that you know the conjunction that the hands point to 4:30 pm and your clock has always worked perfectly reliably. (1)-(3) are all true, but you do not know that it is 4:30 pm any more than you know this in *Stopped Clock*. So Dretske's early analysis is also too weak, predicting knowledge where there is ignorance.

Adams *et al* now object as follows.

> In the case of the Backward Clock, then, one cannot hold the mechanism of the clock fixed in near possible worlds because that circumstance is something that is dependent on the fact that p: that it is 4:30 p.m. That is, there is a dependency relationship between the fact that it is 4:30 p.m. and that the mechanism works the way that it does. Change the mechanism, and you change the time.[67]

This is false. If you wake at 4:30 pm and observe the position of the hands of *Normal Clock*, *Stopped Clock* or *Backward Clock* pointing to 4:30 pm, then the time is still 4:30 pm, despite the fact that mechanisms operate differently. If you wake at 4:15 pm and observe the position of the hands of *Normal Clock*, then you would observe them pointing to 4:15 pm and so would believe that it is 4:15 pm. If you wake at 4:15 pm and observe the position of the hands of *Stopped Clock*, then you would observe them pointing to 4:30 pm and so would believe that it is 4:30 pm. If you wake at 4:15 pm and observe the position of the hands of *Backward Clock*, then you would observe them pointing to 4:45 pm and so would believe that it is 4:45 pm. Nonetheless the time in all three cases would still be 4:15 pm. Indeed it is difficult to even make sense of the idea of changing the time. The time steadily *changes* implacably according to time's arrow. But that cannot *be changed*, unless Adams *et al* countenance time travel.

In my slight modification of *Backward Clock*, you believe that it is 4:30 pm in circumstances in which its mechanism drives its hands reliably backwards from

[67] Clarke, Adams, and Barker, "Methods Matter," 109.

5:00 pm to 4:00 pm. In these circumstances, your clock is *unreliable as an indicator of actual time*. Nonetheless, the position of its hands is deterministically and predictably correlated with the time it indicates. So like *Thermometer*, it is also *reliable in the way its mechanism operates*. These circumstances are logically and causally independent of the fact that it is 4:30 pm. The fact that your clock is running backwards neither entails nor causes it to be 4:30 pm. Conversely, the fact that it is 4:30 pm neither entails nor causes your clock to run backwards.

8. The Appeal to Unfair Internalism

There remain a few other objections to mop up. One of these is as follows

> Williams is not sensitive to the distinction between what the clock DISPLAYS, i.e., the position of its hands, and what the clock SAYS, i.e., what it designedly indicates about the current time, and how that distinction functions in Nozick's account of method and Dretske's account of reasons (in his broad notion of reason as including evidence).[68]

I remain neutral on the question of what *Backward Clock* displays, if anything, and what it says, if anything. All I need to stipulate is that you observe the position of its hands.

Adams *et al* also complain that I am unfairly internalist. They say the following.

> Williams makes his first mistake here by construing Dretske's notion of a reason as referring only to a premise in an argument from an internalist perspective. Dretske does not restrict the idea of a reason in that way, but includes one's evidence, i.e., facts one knows to obtain, even if one is not aware of what one's evidence is.[69]

I did not restrict it that way either. The fact is that the hands of *Backward Clock*, one that has always worked perfectly reliably as *Normal Clock*, now point to 4:30 pm. You might use this fact as a premise in an argument. On the other hand, it might constitute evidence that you use pre-reflectively and unconsciously.

Adams *et al* also say the following.

> It should also be noted that Williams talks about your **knowing** that the clock has always been reliable in the Backward Clock case, but that is the kind of **internalist** talk that externalists eschew.[70]

[68] Clarke, Adams, and Barker, "Methods Matter," 110-111.
[69] Clarke, Adams, and Barker, "Methods Matter," 106-107.
[70] Clarke, Adams, and Barker, "Methods Matter," 111.

If we follow Timothy Williamson in holding that "knowledge, and only knowledge, constitutes evidence,"[71] then Adams *et al* already have what they want. But in any case my stipulation that you know that your clock has always been reliable appears in my example as part of a *challenge* to a particular variety of externalism. In so far as that stipulation is internalist and helps to derail the tracking theories, it contributes to a victory for internalism over externalism. That is not grounds for complaint.

We should also note that your method of ascertaining the time, namely by observing the position of the hands of your clock during the hour that you nap, is one that you may use pre-reflectively and unconsciously. You may use it, as Nozick puts it, "from the inside."[72] From your point of view, things are exactly the same at 4:30 pm whether you are in *Normal Clock*, *Stopped Clock* or *Backward Clock*.

9. Concluding Remarks

In their landmark 2005 paper, Adams and Clark mounted an impressive defence of the tracking theories against putative counterexamples. Adams, Barker and Clark have now put up a staunch defence of the tracking theories against *Backward Clock*. This fails. Appeals to sensitive or reliable methods or to extra-sensitivity do not help them. They are still stuck with the fact that the tracking theories fall to *Backward Clock*, an even more useful test case for other analyses of knowledge than might have first appeared[73].

[71] Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford. University Press, 2000), 185.
[72] Nozick, *Philosophical Explanations*, 185.
[73] I am grateful to David Blaxton for suggesting the title of this paper.

REVIEWS

Tobias Endres, Pelegrinno Favizzi, Timo Klattenhoff (editors), *Philosophie der Kultur- und Wissensformen. Ernst Cassirer neu lesen*
(Frankfurt am Main, Bern, Brussels, New York, Oxford, Warszawa, Vienna: Peter Lang, 2016)

Reviewed by Ioan Alexandru Tofan

Revisiting a classic such as Ernst Cassirer is far from easy. Firstly, the most recent trends in philosophy indicate a propensity for deconstruction and experiment rather than re-reading traditional texts or restoring great authors of the past. Secondly, at a first glance, Cassirer himself seems to resist re-reading: his precise distinctions, the attention and rigour of this discourse, as well as the vast range of interests (his complete works, published by Meiner Verlag, stretch over 26 volumes, and his posthumous edition just 18 for the time being), sometimes daunts the reader who may hesitate upon testing his concepts by taking a fresh line of questioning away from the trodden path.

And yet, the book edited by Tobias Endres, Pelegrinno Favizzi and Timo Klattenhoff is a certain success. Its contributors are not "hurried" readers. As members or admirers of the *Cassirer-Arbeitsgruppe* (Technische Universität, Berlin) and doctoral students of Cassirer's work, they start from the premise that, read carefully, the philosopher's concepts are already open to perpetual reassessment, even a transdisciplinary reassessment that constantly pushes the pre-established borders between sciences or the boundaries of some phenomenological analysis: "So bewegt sich die Philosophie Cassirers programmatisch in einer Zwischensphäre, in einem infinitesimalen Raum zwischen den vielfältigen Ausdrucksmöglichkeiten des menschlichen Geistes: Sie zeichnet sich insofern durch einen ausgeprägt transdisziplinären Charakter aus, als dass sie die Wissensfelder nicht als vorgefertigte Gebiete annimmt, sondern sie integrativ anspricht, um Vernetzung und Interaktion zu fördern und dadurch die Schaffung neuer Erkenntnis zu ermöglichen" (p. 13). The philosophy of symbolic forms is mainly a transcendental enterprise permanently open to re-evaluation, not a closed conceptual construct. Christian Möckel's article (*Symbolische Formen als Wissensformen?*) highlights the plasticity of Cassirer's philosophical terms and the great potential for assuming them in phenomenological and anthropological contexts. Equally interesting is the discussion about the way in which Cassirer re-interprets hallmarks of philosophical tradition (such as the Hegelian dialectics explored by Sevilay Karaduman), or the connections he makes with contemporary

issues of the action theory (via Joel-Philipp Krohn's discussion of American pragmatism). In other words, a renewed reading of the German philosopher's work defines a series of models in Cassirer's very *modus operandi* when he formulates his own thinking and way of looking at the world.

But a competent reading is not enough. The reader needs to take the appropriate perspective to Cassirer's concepts. A philosophy's topicality may be understood in different ways and, on many occasions, unabashed persistence may even end up twisting it in the wrong direction turning it into a token of academic performance in the field of humanities. But the authors point to an important aspect that saves them from idle sensationalism: that the *Cassirer-Reinassance* of the nineties is in fact an internationalization of his work (the editors talk about a globalized Cassirer). As the Contents page shows, basically the entire international academia has undertaken a "new" reading of the philosophy of symbolic forms. So, in this context, "new" means 'different' rather than 'recent,' marked by a multitude of social and cultural realities. How can a series of concepts derived from the essence of Western tradition be applied beyond the phenomenological field it is faced with? In what way can new worlds such as the Internet (as in Rafael Garcia's text) or contemporary cinematography (as in Peter Remmers' article) bear Cassirer's analysis of cultural forms? The views around classic instances of symbolic forms (such as painting in Yosuke Hamada's analysis of aesthetic intuition, or money in Timo Klattenhoff's parallel reading of Cassirer and Simmel) converge around a practically infinite universe of symbolic forms – forms of a surprising complexity and historical evolution. A second goal of reading, rooted in Cassirer's own interests but still open to unlimited reformulation, is of political nature in the wide sense of symbolic construct of human reality. In this context, Pellegrino Favuzzi explores the possibility to integrate reason and emotion in the definition of a "rational pathos" (*Vernunftpathos*) of *zoon politikón*, and Gisela Starke studies the mythological structure of National Socialism in order to decipher its totalitarian mechanisms. As to the reference to Cassirer, the novelty is a systematic appeal for plurality and tolerance via the critique of the redefined forms of culture and of its reception. As Servanne Jollivet places Cassirer's writings in the *Historismus-Debatte* context, she studies the relativism/dogmatism dualism that dominates the discussion of man's historical situation, and she develops the idea of a dynamic unity with immediate consequences in the way we perceive, and talk to, the *other*.

The issues regarding reassessing some basic principles of Cassirer's vision (such as the relationship between the philosophy of symbolic forms and an integrative theory of perception – in Tobias Endres' text; the question whether

philosophy itself can be seen as a symbolic form – with Claudio Bonaldi; or Felix Schwartz's critical discussion on Cassirer's anti-Naturalism) complete the series of the above-mentioned specific topics. Thus, the contemporary re-evaluation of Cassirer's work is not restricted to highlighting the relevance of his concepts in various current theoretical disputes, but implies the effort of re-reading him in depth and dynamically understanding him in the fundamental premises of his discourse. It is an organic, integrative perspective on the method and perspective that Cassirer applies to culture, politics, or the stake and forms of knowledge, and it ensures consistency and coherence to any effort of re-reading and interpretation. "Ernst Cassirer neu lesen" becomes „Ernst Cassirer neu begegnen und zusprechen."

The success of this book's enterprise is also due to the fact that it is the result of a lively dialogue between the contributors. Dr. Martina Plümacher and Dr. Christian Möckel are founding members and catalysts of the Ernst Cassirer-Arbeitsgruppe as part of the *Innovationszentrum Wissensforschung der Technischen Universität* Berlin. The debates within this study group featured some converging topics coming from a variety of research areas. All of them meet around Cassirer's figure seen in a fresh light by today's readers.

NOTES ON THE CONTRIBUTORS

Miloud Belkoniene is a PhD candidate at Fribourg University, Switzerland. His main areas of interest are epistemology, philosophy of mind, philosophy of science and philosophy of action. His current work focuses on the distinction that can be drawn between probabilistic credence and rational belief in relation to the features of evidence to which each of these states is sensitive and to their role in the analysis of rational decisions. Contact: miloud.belkoniene@unifr.ch.

John Biro teaches at the University of Florida. His interests are in the philosophy of language, the philosophy of mind, epistemology, metaphysics and the history of philosophy, especially in the seventeenth and eighteenth centuries. He has published papers in these areas in journals such as *The Monist, Australasian Journal of Philosophy, Philosophy and Phenomenological Research, Philosophical Studies, History of Philosophy Quarterly, The Review of Metaphysics and Hume Studies.* He is also co-editor of two collections of papers on Spinoza and one on Frege. Contact: jbiro@ufl.edu.

Fabio Lampert is a philosophy graduate student at the University of California, Davis. His main interests are in philosophical logic, logic, modality, epistemology, philosophy of language, and philosophy of religion. His publications include "Actuality and the A Priori" (*Philosophical Studies*, 2017) and "Actuality, Tableaux, and Two-Dimensional Modal Logics" (*Erkenntnis*, 2017). Contact: flampert@ucdavis.edu.

Andy Mueller is research assistant at J.W. Goethe University Frankfurt. He received his PhD from Goethe University in 2016 with a thesis on pragmatic encroachment and epistemic norms for practical reasoning. He is mainly interested in epistemology and in all things concerning reasons and rationality, both epistemic and practical. His paper "How Does Epistemic Rationality Constrain Practical Rationality?" and his co-authored paper "Knowledge Dethroned" (with Jacob Ross) are both forthcoming in *Analytic Philosophy*. Contact: andreas.mueller.uniffm@gmail.com.

Michael J. Shaffer is a professor of philosophy at St. Cloud State University. He has published numerous articles in the philosophy of science, logic and epistemology. He is the co-editor of *What Place for the A Priori?* (Chicago: Open Court, 2011) and author of *Counterfactuals and Scientific Realism* (Palgrave-MacMillan, 2012) and *The Experimental Turn and the Methods of Philosophy (*New York: Routledge, forthcoming). Contact: shaffermphil@hotmail.com.

Ioan Alexandru Tofan teaches Philosophy of Religion, Romanian Philosophy and Critical Theory at the "Alexandru Ioan Cuza" University of Iasi, Romania. He wrote books and articles on Hegel, Walter Benjamin, Michel de Certeau. He currently writes a book on the concept of spiritual experience in the works of Andre Scrima. Contact: atofanro@yahoo.com.

John N. Williams is an associate professor in philosophy in the School of Social Sciences at Singapore Management University. His research interests include paradoxes, epistemology and applied ethics. His most recent articles include: "There's Nothing to Beat a Backward Clock" (*Logos & Episteme*, 2016), "The Backward Clock, Truth-Tracking, and Safety" (with Neil Sinhababu, *Journal of Philosophy*, 2015), "Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge" (*Analysis*, 2015), "Moore's Paradox in Speech: A Critical Survey" (*Philosophy Compass*, 2015), and "Moore's Paradox in Thought: A Critical Survey" (*Philosophy Compass*, 2015). Contact: johnwilliams@smu.edu.sg.

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 take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer.
The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.
The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.

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. Property & Royalties

Articles accepted for publication will become the property of *Logos & Episteme* and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Permissions

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

15. 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.