

**Volume XIV** ♦ **Issue 3**

2023

# Logos & Episteme

an international journal  
of epistemology

**Romanian Academy  
Iasi Branch**



**“Gheorghe Zane” Institute  
for Economic and Social  
Research**

## **Founding Editor**

---

Teodor Dima (1939-2019)

## **Editorial Board**

---

### **Editor-in-Chief**

Eugen Huzum

### **Executive Editors**

Vasile Pleșca  
Cătălina-Daniela Răducu

### **Assistant Editors**

Irina Frasin  
Bogdan Ștefanachi  
Ioan Alexandru Tofan

### **Web&Graphics**

Codrin Dinu Vasiliu  
Virgil-Constantin Fătu  
Simona-Roxana Ulman

## **Contact address:**

---

Institutul de Cercetări  
Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

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

[https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc\\_logos-episteme](https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme)

## **Advisory Board**

---

Frederick R. Adams  
University of Delaware, USA

Scott F. Aikin  
Vanderbilt University, USA

Daniel Andler  
Université Paris-Sorbonne, Paris IV, France

Panayot Butchvarov  
University of Iowa, USA

Mircea Dumitru  
Universitatea din București, România

Sanford Goldberg  
Northwestern University, Evanston, USA

Alvin I. Goldman  
Rutgers, The State University of New Jersey, USA

Susan Haack  
University of Miami, USA

Stephen Hetherington  
The University of New South Wales, Sydney, Australia

Paul Humphreys  
University of Virginia, USA

Jonathan L. Kvanvig  
Baylor University, USA

Thierry Martin  
Université de Franche-Comté, Besançon, France

Jürgen Mittelstrab  
Universität Konstanz, Germany

Christian Möckel  
Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot  
Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins  
Universidade de Lisboa, Portugal

Duncan Pritchard  
University of Edinburgh, United Kingdom

Nicolas Rescher  
University of Pittsburgh, USA

Rahman Shahid  
Université Lille 3, France

Ernest Sosa  
Rutgers University, USA

John F. Symons  
University of Texas at El Paso, USA

# TABLE OF CONTENTS

## RESEARCH ARTICLES

Frederik J. ANDERSEN, Uniqueness and Logical Disagreement (Revisited).....	243
Arnold CUSMARIU, Is JTB Knowledge Hopeless?.....	261
Danilo Fraga DANTAS, When the (Bayesian) Ideal Is Not Ideal.....	271
Erhan DEMIRCIOGLU, Subjective Rationality and the Reasoning Argument....	299
Andreas STEPHENS, Contextual Shifts and Gradable Knowledge.....	323
Notes on the Contributors.....	339
Notes to Contributors.....	341
<i>Logos and Episteme</i> . Aims and Scope.....	345



## **RESEARCH ARTICLES**



# UNIQUENESS AND LOGICAL DISAGREEMENT (REVISITED)

Frederik J. ANDERSEN

ABSTRACT: This paper discusses the *Uniqueness Thesis*, a core thesis in the epistemology of disagreement. After presenting uniqueness and clarifying relevant terms, a novel counterexample to the thesis will be introduced. This counterexample involves *logical disagreement*. Several objections to the counterexample are then considered, and it is argued that the best responses to the counterexample all undermine the initial motivation for uniqueness.

KEYWORDS: the uniqueness thesis, rational uniqueness, logical disagreement, logical evidence, propositional justification, epistemic permissivism, peer disagreement

## 1. Introduction

The *Uniqueness Thesis* (henceforth denoted ‘UT’) concerns a relation between a body of evidence, a doxastic attitude, and a proposition.<sup>1</sup> Jonathan Matheson, a proponent of the thesis, defines UT as follows:

(UT) For any body of evidence  $E$  and proposition  $[p]$ ,  $E$  justifies at most one doxastic attitude toward  $[p]$  (Matheson 2011, 360).

UT features frequently in the epistemology literature,<sup>2</sup> especially in the debate concerning peer disagreement—if two epistemic peers<sup>3</sup> disagree about a proposition  $p$ , is it then possible that they are both justified in their doxastic attitudes toward  $p$ ? If UT is true, the answer is negative.

Importantly, there are in fact several non-equivalent definitions of UT in the literature. Thomas Kelly, for example, favors a formulation of UT saying that there

---

<sup>1</sup> This paper is largely based on Andersen (2020), but it includes several important corrections and additions.

<sup>2</sup> See for example Matheson (2011), Rosa (2012, 2016), Kelly (2014), White (2014), Kopec and Titelbaum (2016), Ross (2021), Kauss (2023).

<sup>3</sup> Roughly put, two agents in disagreement are epistemic peers when neither side is epistemically superior with respect to the target-proposition at hand, i.e., when the two are similar enough in all relevant factors such as evidence, track record, time constraints etc.

is *exactly one* justified doxastic attitude given a body of evidence (Kelly 2010, 119), while Matheson prefers *at most one*, as we have just seen. Matheson notes that in most cases there will be exactly one justified doxastic attitude given a body of evidence, but in some situations, there may be no justified doxastic attitude toward  $p$  whatsoever. This can arguably happen when one is not able to, or when it is simply not possible to, comprehend the proposition at hand.<sup>4</sup> If one takes (possible) comprehension of  $p$  to be a necessary condition for the existence of a justified doxastic attitude toward  $p$ , then it seems most reasonable to use Matheson's weaker definition of UT. Thus, this is what we will assume here.

Further, we will adopt Matheson's assumption that the term 'doxastic attitude' can only refer to the following three possibilities: *belief that  $p$* ; *disbelief that  $p$* ; and *suspension of judgement with respect to  $p$* ; i.e., the possibility space of attitudes that one can take toward a proposition  $p$  is exhausted by these three attitudes.<sup>5</sup>

Now, UT puts a constraint on the total number of doxastic attitudes that a body of evidence can justify toward a proposition. According to UT any body of evidence  $E$  justifies at most one doxastic attitude toward  $p$ . In other words, according to UT, there exists no body of evidence  $E$  such that  $E$  justifies both belief and disbelief toward  $p$ . Similarly, of course, the thesis implies that there exists no  $E$  such that  $E$  justifies both a (dis)belief in  $p$  and suspension of judgement with respect to  $p$ . In the paper titled '*The case for Rational Uniqueness*', Matheson makes two further clarifying remarks about UT:

(UT) [...] makes no reference to individuals or times since (UT) claims (in part) that who possesses the body of evidence, as well as when it is possessed, makes no difference regarding which doxastic attitude is justified (if any) toward any particular proposition by that body of evidence (Matheson 2011, 360).<sup>6</sup>

---

<sup>4</sup> See Feldman (2006) for a motivation of this view.

<sup>5</sup> This assumption is common in the contemporary literature, see for example Kelly (2010), Matheson (2011), Rosa (2012), Titelbaum (2015, 2019). Note that some have argued that the doxastic attitude of *disbelief that  $p$*  is non-equivalent to that of *believing the negation of  $p$* . See Smart (2021) for a recent argument. Unless otherwise stated we'll simply take disbelief that  $p$  and believing the negation of  $p$  as equivalent attitudes in what follows.

<sup>6</sup> Note that while Matheson's statement of UT doesn't make reference to individuals (i.e., cognizers or human agents) at all, some authors have presented versions of uniqueness that do. Consider for example Titelbaum and Kopec's tripartite distinction between propositional, attitudinal, and personal uniqueness (Titelbaum and Kopec 2019, 206). *Propositional Uniqueness*. Given any body of evidence and proposition, the evidence all-things-considered justifies either the proposition, its negation, or neither. *Attitudinal Uniqueness*. Given any body of evidence and proposition, the evidence all-things considered justifies at most one of the following attitudes



(UT) concerns propositional justification, rather than doxastic justification. That is, the kind of justification relevant to (UT) is solely a relation between a body of evidence, a doxastic attitude, and a proposition. How individuals have come to have the doxastic attitudes they have toward the proposition in question will not be relevant to our discussion. Further, individuals can be propositionally justified in adopting attitudes toward propositions which they psychologically cannot adopt [...] Importantly, it is not a necessary condition for being justified in believing [ $p$ ] that one be able to demonstrate that one is justified in believing (Matheson 2011, 360-361).

The first of these quotes states that according to UT a given body of evidence  $E$  justifies exactly the same doxastic attitude (if any) towards  $p$ , no matter the subject that assesses  $E$  and at what time this is done. In the second quote, Matheson distinguishes between *propositional* and *doxastic* justification, where the former is a relation between a body of evidence, a doxastic attitude, and a proposition, the latter concerns *how* a given individual came to adopt a specific doxastic attitude towards a proposition, i.e., doxastic justification is concerned with one's reasons for actually adopting a certain attitude toward  $p$ . Doxastic justification presumes that a given individual has a certain attitude toward  $p$ , and the question is then whether or not this individual has sufficiently good (epistemic) reasons to be justified in having that attitude.<sup>7</sup> When it comes to propositional justification, on the other hand, it is irrelevant whether any individual is ever concerned with  $p$ ; the crux of propositional justification is that a justification-relation between a body of evidence, a doxastic attitude, and a proposition holds, not whether any individual realizes this. Understood in this way propositional justification refers to an external relation, and an individual can accordingly be propositionally justified in a doxastic attitude towards  $p$  even though this individual has not adopted the relevant attitude psychologically. And hence, it is not necessary for a subject to be able to demonstrate or defend this given attitude towards  $p$  in order for it to be propositionally justified. Matheson tells us that UT is a thesis concerning propositional justification rather than doxastic justification.

## 2. Clarifications

---

toward the proposition: belief, disbelief, or suspension. *Personal Uniqueness*. Given any body of evidence and proposition, there is at most one doxastic attitude that any agent with that total evidence is rationally permitted to take toward the proposition.

<sup>7</sup> For accounts of the epistemic *basing relation*, which is often taken to be relevant for doxastic justification, see for instance McCain (2014), Carter and Bondy (2019), Korcz (2021).

Before we move on to consider the announced counterexample to UT, let us pause to further specify what is meant by ‘justification’ and ‘evidence’ in the rest of the paper. We will deliberately stay on a high level of generality in order not to exclude too many accounts of justification and evidence from the later discussions in sections 3 and 4.

When using the term ‘justification,’ this use is restricted to the epistemic domain, we are not concerned with any practical issues whatsoever. So, in other words, our concern is with the justification of doxastic attitudes towards propositions. This kind of justification is taken to be regulated by epistemic norms, i.e., truth-conducive norms, and as indicated in §1, we are concerned with *propositional* justification rather than doxastic justification.<sup>8</sup>

Our use of the term ‘evidence’ assumes that we can all agree that evidence can stem from many different sources like direct visual perception, testimony from individuals or media, scientific experiments etc. The only constraints we will force on our understanding of evidence from the outset are: (1) evidence must be propositional (and thus truth-apt); (2) any piece of evidence must be true; (3) any piece of evidence must (at least in principle) be accessible to human beings; and (4) evidence should be supportive of doxastic attitudes, where ‘support’ may be interpreted probabilistically, but does not have to be.

(2) is arguably the most controversial among these four constraints. For our purposes, however, there is a very good reason for including this factivity condition. To see this, suppose that one could have false pieces of evidence in one’s (total) body of evidence  $E$ . Then, given the further assumption that false evidence can support anything, we could easily have a situation where a true bit of evidence  $e_1$  from  $E$  entails  $p$  and thus supports the belief that  $p$ , while a false bit of evidence  $e_2$  from  $E$  entails  $\text{not-}p$  and thus supports disbelieving that  $p$ , making  $E$  inconsistent and “explosive.” This would in effect trivialize the debate about UT; on this account of evidence UT is obviously false.<sup>9</sup> Hence, we should either accept that evidence is factive or we should deny that false evidence can support anything. For the rest of the paper we will take the first option.

### 3. The Argument from Logical Disagreement

---

<sup>8</sup> The literature on epistemic justification is vast, but prominent examples of theories of justification can be found in Goldman (1986), Bonjour (1985), Feldman and Conee (1985), Alston (1989), Williamson (2000), Conee and Feldman (2004). Note also Littlejohn’s tripartite division of epistemic justification which includes *personal* justification as well as doxastic and propositional (Littlejohn 2012, 5). According to Littlejohn, doxastic justification is sufficient for personal justification, but not *vice versa*.

<sup>9</sup> Thanks to Franz Berto for pressing this point about false evidence.

Consider now the following case against UT:

**Logical Disagreement.** Two logicians,  $S_1$  and  $S_2$ , are walking into an empty auditorium where they find a deduction written on a blackboard.  $S_1$  and  $S_2$  are simultaneously looking at the board. As it happens,  $S_1$  is a classical logician, while  $S_2$  is an intuitionist. Now, by definition, the deduction consists in a finite number of steps, so all steps of the deduction except for the conclusion  $C$  will serve as a common body of evidence  $E$ , i.e., a set of propositions that are represented in a language that both logicians fully comprehend. The central question is then whether  $E$  entails  $C$ . Suppose that conclusion  $C$  on line  $n$  is the result of applying DNE (double negation elimination) to not-not- $C$  on line  $n - 1$ .<sup>10</sup> As  $S_1$  accepts classical logic, she also accepts the inference from not-not- $C$  to  $C$ , while  $S_2$  given her intuitionist convictions denies DNE as a general rule of inference and thus denies that  $C$  needs to come out supported by  $E$ .

In this case we have a situation in which two agents possess exactly (!) the same evidence (the propositions represented by lines  $n - 1$  on the blackboard), but they are justified in diverging doxastic attitudes toward the relevant proposition in question, namely  $C$ . We see that  $E$  justifies  $S_1$  in her belief that  $C$ , while  $E$  justifies (at least) suspension of judgement regarding  $C$  for  $S_2$  (as  $C$  is not necessarily supported by  $E$ ). Thus, the case is a clear counterexample to UT as the number of attitudes that  $E$  justifies exceeds one.

Of course, as the reader will have noticed by now, the case is concerned with a special type of evidence, i.e., evidence of the completely formal type that we find in pure logic and mathematics. This means that the counterexample is narrow in the sense that it does not indicate the existence of counterexamples to UT among other types of evidence.<sup>11</sup> However, this will be completely irrelevant as

---

<sup>10</sup> Using standard notation that isn't meant to favor any logical tradition, DNE is an inference from  $\Gamma \vdash \sim\sim\varphi$  to  $\Gamma \vdash \varphi$ , where ' $\Gamma$ ' denotes a set of sentences in a given language, ' $\vdash$ ' denotes deducibility from left to right, ' $\sim$ ' denotes a negation operator, and ' $\varphi$ ' picks out a single sentence of the language. Some readers may point out that it is underspecified in the case above whether  $S_1$  and  $S_2$  disagree over an *instance* or a *schema* of DNE. This is true, but it will not make a significant difference to the main argument of the paper. The crux is that the logicians genuinely disagree. For more elaborate discussions of genuine logical disagreement the reader should consult Hattiangadi (2018), Hjortland (2022), Andersen (2023b), Hattiangadi and Andersen (202X).

<sup>11</sup> However, some epistemologists have suggested that there are counterexamples to UT among other types of evidence. Consider, for example, a case where  $S_1$  and  $S_2$  discuss which football team will win the national league this season. Suppose that their discussion takes place the day before the final match day, and at this point of the season only two teams can win; either team A or team B (not both). Suppose further that the only evidence available to the subjects is a certain newspaper statistic, which shows the scores of the season so far. According to this statistic, team A is in front of team B by the smallest possible margin. Now,  $S_1$  is convinced that team A will

long as we regard UT as a *general* epistemic principle. If the case holds, we will have a counterexample sufficient for rejecting UT.

Finally, before taking on some pressing objections to the Argument from Logical Disagreement, one further clarifying comment is called for. Note that the logical disagreement described above isn't simply a case where  $S_1$  and  $S_2$  are talking past each other because of equivocation about the meaning of the expression 'not', as Quine (1986, 81) would have it. The reason why we can rule this out is a certain "technique for arguing that an apparent conflict is a real one" due to Williamson (1988).<sup>12</sup> In (1982) Harris established that in a system of natural deduction with two different operators for negation—classical (' $\neg$ ') and intuitionist (' $\rightarrow$ '), respectively—the biconditional  $\neg\varphi \leftrightarrow \rightarrow\varphi$  becomes provable, for any formula  $\varphi$ . From this basis Williamson's technique requires us to ask whether (i) there are rules of inference governing both  $\neg$  and  $\rightarrow$ , and (ii) whether such rules could allow classical and intuitionist logicians (like  $S_1$  and  $S_2$ ) to characterize negation as the unique operator obeying those rules (up to logical equivalence).

As it turns out, the answer to (i) is positive: both  $\neg$  and  $\rightarrow$  obey Ex Falso Quodlibet (' $EFQ$ ') and the Introduction Rule for Negation (' $N_{Intro}$ '). Let  $\varphi, \psi$  be well-formed formulas. Then a monadic operator  $\sim$  obeys  $EFQ$ ,  $N_{Intro}$ , and  $N_{Elim}$ , just in case the following two schemas are valid:

$$\frac{\varphi \quad \sim\varphi}{\psi} \text{ } EFQ \qquad \begin{array}{c} \overset{(n)}{\varphi} \\ \vdots \\ \frac{\perp}{\sim\varphi} \end{array} \text{ } \overset{(n)}{N_{Intro}}$$

---

take the championship due to the statistical support for this (they are ahead at this point). However,  $S_2$  suspends judgement about who will be the champions as team A leads with the smallest possible margin and it is still possible for team B to make it. In such a case the proponent of UT should say that at most one of the subjects' doxastic attitudes is justified, but one might argue that this is wrong. In such borderline cases it may seem that at least two out of three doxastic attitudes could be justified. If this is right, we have a counterexample to UT featuring another type of evidence, i.e., empirical data. Find similar borderline cases in (Kelly 2014, 299-300). For a recent discussion of (merely) statistical evidence and its role in epistemology, see Silva (2023).

<sup>12</sup> Note that our exhibition of Williamson's technique follows the order of presentation found in Rossi (2023). We follow Rossi's lead as his presentation of the material is very clear and detailed.

Here, numerals in brackets, i.e.,  $(n)$ , serve two distinct purposes: they mark discharged assumptions; and they indicate at which point in the derivation assumptions are discharged.

The answer to (ii) is also positive.  $EFQ$  and  $N_{Intro}$  are jointwise strong enough to define any monadic operator obeying them (up to logical equivalence). To see this, let  $\sim_1$  and  $\sim_2$  be any two monadic operators obeying  $EFQ$  and  $N_{Intro}$ . The following derivation establishes the deductive equivalence:  $\vdash \sim_1 p \leftrightarrow \sim_2 p$ .

$$\frac{\frac{\frac{(1) p}{p} \quad \frac{(2) \sim_1 p}{\sim_1 p} \quad EFQ}{p} \quad \frac{(1) p \quad (2) \sim_1 p}{\sim_2 p} \quad EFQ}{\sim_1 p \rightarrow \sim_2 p} \quad (2) \rightarrow_{Intro} \quad \frac{\frac{(3) p}{p} \quad \frac{(4) \sim_2 p}{\sim_2 p} \quad EFQ}{p} \quad \frac{(3) p \quad (4) \sim_2 p}{\sim_1 p} \quad EFQ}{\sim_2 p \rightarrow \sim_1 p} \quad (4) \rightarrow_{Intro}}{\sim_1 p \leftrightarrow \sim_2 p} \quad \leftrightarrow I$$

As the answers to both (i) and (ii) are positive, Williamson (1988, 111) proposes a proof-theoretic argument showing that the disagreement between classical and intuitionist logicians over DNE is a genuine one, and not merely a verbal dispute. Summa: If there is only one monadic operator—up to logical equivalence—obeying both  $EFQ$  and  $N_{Intro}$ , then this must rule out the possibility that the classical and intuitionist logicians are merely talking past each other when disagreeing about whether it obeys DNE. Either the intuitionist is right and the classicist wrong (or *vice versa*). In any case, there cannot be a single logic with two negation operators only one of which obeys DNE.

#### 4. Objections and Responses

As the case presented above will be very hard to accept for many readers (for various reasons), the rest of the paper aims to motivate the Argument from Logical Disagreement. The strategy here is simple. While discussing various objections to **Logical Disagreement**, it will become clear that the UT-proponent can only avoid the counterexample by undermining the initial motivation behind UT, i.e., explaining away the counterexample to UT will lead to an indirect defeat of the thesis. In the following, five objections to **Logical Disagreement** will be scrutinized (§§4.1-4.5). The first two will simply be rejected, the third will be found underdeveloped, and while the remaining two can actually explain away the counterexample to UT, this can only be done by undermining the motivation behind the principle.

#### 4.1 Evidence is Contingent

**Objection 1.** Even though the evidence  $E$  present in **Logical Disagreement** satisfies our four rudimentary constraints on evidence (cf. §2) as  $E$  is propositional, factive, accessible, and supportive,  $E$  is still not a genuine body of evidence. For only contingent propositions can be evidence. Thus, UT is not even applicable in **Logical Disagreement**.

First of all, there is no principle reason why necessary propositions such as the ones found in pure mathematics and logic cannot be counted as evidence. Propositions of logic and mathematics can clearly serve the supportive role of evidence very well, i.e., such propositions speak in favor of certain hypotheses in the strongest possible way (by entailment). Hence, if any proposition is able to justify a belief, it seems that pure logical or mathematical propositions are ideal candidates. Habit may dictate, perhaps leading back to acceptance of Hume's Fork, that some of us cannot see the point in taking purely formal premises of deductive arguments as evidence, but without further qualification this is obviously not a good argument for accepting such an exclusion in philosophical or scientific work. Moreover, accepting **Objection 1** leads to absurd consequences when we hold other plausible epistemic principles to be true. Take for example Timothy Williamson's principle  $E = K$ , i.e., evidence equals knowledge (2000, Chapter 9). If we accept that our evidence is coextensive with our knowledge, and that **Objection 1** holds, it directly follows that we cannot have pure mathematical or logical knowledge. To deny that we can and do have such knowledge would not only be absurd, it would be intellectual suicide.

#### 4.2 Communication Breakdown

**Objection 2.** The case **Logical Disagreement** misrepresents the interaction between classical logicians and intuitionists. Where the classical logician works with a philosophical presupposition of a realm of mathematical objects independent of the thinking subject (objects that obey the laws of classical logic and can stand in set-theoretic relations), this is radically different from the intuitionists who advocate for constructive methods and take mathematics to be about mental constructions. As a result of this schism, the two logicians in the proposed case would run into an insurmountable communication breakdown, i.e., the DNE-inference acceptable to the classical logician would not even be understandable to the intuitionist—it would be nonsense. To quote Brouwer: "*Let us now consider the concept: 'denumerably infinite ordinal number.' From the fact that this concept has a clear and well-defined meaning for both formalist and intuitionist, the former infers the right to create the 'set of all denumerably infinite ordinal numbers,' the power of which he calls aleph-one, a right not recognized by the intuitionist.*" (Brouwer, 1975). Something similar to what

Brouwer describes in the interaction between diverse traditions in this quote occurs in **Logical Disagreement** with respect to DNE, i.e., the intuitionist does simply not comprehend the final step of the deduction on the blackboard. Thus, suspension of judgement is not a justified doxastic attitude for the intuitionist in this case; the supposed logical connection between *E* and *C* is gibberish to her. Rather, **Logical Disagreement** represents the kind of case where there is no justified doxastic attitude for the intuitionist to have. Hence, UT would be saved (at least the *at most one doxastic attitude*-version of the thesis). The case allows only one justified attitude, namely the attitude of the classical logician.

This objection overstates the divide between the classical and intuitionist traditions. Comprehension of classical logic is often presupposed in discussions of non-classical logical systems, e.g., as a metatheory. Indeed, it is stipulated in **Logical Disagreement** that the deduction found on the blackboard is written in a language that both logicians fully comprehend. We do not need more than noticing and appreciating this very stipulation in order to slide off the objection.

Further, we can strengthen this reply by noticing that it is not the case that when there is logical disagreement, one party has automatically misunderstood (or lacks) some concept. The disagreement may just be the result of one side having false beliefs. So, in **Logical Disagreement**, it need not be the case that the intuitionist (supposing that she got it wrong) lacks some concept about how negation works, or has misunderstood or changed its meaning. Negation means whatever it means, also in the intuitionist's mouth, she just has false beliefs about that meaning.<sup>13</sup>

### 4.3 Logical Monism

Now, let us turn to the more challenging objections.

**Objection 3.** The evidence *E* does in fact justify exactly one doxastic attitude in **Logical Disagreement**, it is just that we do not know which attitude it is. For we do not know which logical theory is the "correct" model of logical consequence, but surely there is only one correct logic in the end. Thus, UT survives the case even though the logical disagreement between the classical logician and intuitionist leaves us in the dark with respect to which doxastic attitude is justified by *E*.<sup>14</sup>

This objection begs the question against *logical pluralists* (e.g., Beall & Restall-style), i.e., the view that there is more than one true (or correct) logic.<sup>15</sup> According

---

<sup>13</sup> A similar point is made by Williamson in (2007, Chapter 4).

<sup>14</sup> See, e.g., Griffiths and Paseau (2022) for a recent defense of *logical monism*.

<sup>15</sup> In principle, the objection also begs the question against *logical nihilism*, which is the extreme view that there is no true (or correct) logic at all (Russell 2018).

to logical pluralists, there is not always a single answer to the question whether a proposition  $p$  logically follows from a set of propositions (premises), in some cases there are more than one correct answer. A rough motivation for logical pluralism is that theories of classical logic, relevance logic, intuitionistic logic etc., all have a rightful place in formalizing and restraining logical inference as various important aspects of our pre-theoretic notion of logical consequence can be explicated by each of these approaches to logic.

Clearly, begging the question against the pluralist in this way merely relocates the tension from an infight between UT-supporters and -deniers to a clash between logical monism and pluralism, so it seems like a dissatisfying option. Of course, some UT-supporters might be happy to say that logical pluralism is false, and thus they will have a way to save their principle, but this strategy should be supported by strong independent reasons. It will not be enough for the UT-supporter to accept logical monism because it seems like the default position amongst mainstream epistemologists. Hence, **Objection 3** is underdeveloped as it stands, and UT-supporters opting for this way out have further work to do.

Developing the back and forth between logical monists and pluralists any further here would take us beyond the scope of this paper, but the reader can find some useful references in the footnote below.<sup>16</sup>

#### 4.4 Splitting the Evidence

**Objection 4.** As  $S_1$  and  $S_2$  belong to two opposing traditions in logic and don't accept the same rules of inference, it is actually not the case that they possess the same evidence in the situation described. Surely, considered just as a set of (formal) propositions, the evidence is the same for both subjects, but due to the subjects' diverse logical backgrounds the evidence splits in two. The case really presents both  $E$  and  $E^*$ , where the acceptable inference rules of classical logic are tacitly accepted to induce  $E$  and the rules of intuitionist logic are tacitly accepted to induce  $E^*$ . No set of (formal) propositions supports anything pre-theoretically. Choosing a logical theory is necessary to even generate *logical* evidence. Pre-theoretically, the question of which doxastic attitude is supported by a body of logical evidence is empty. Hence, **Logical Disagreement** is not a counterexample to UT since each body of evidence only justifies one doxastic attitude.

*Prima facie*, this objection seems to have something going for it. Indeed, it might save UT seen as a general epistemic principle since at most one doxastic attitude can be justified per body of evidence. However, at the same time it undermines the

---

<sup>16</sup> For more on logical pluralism in the Beall & Restall-style, see, e.g., (Beall and Restall, 2000, 2006). Other kinds of logical pluralism can be found in Carnap (2014), Shapiro (2014). For an extensive overview, see Russell (2019).



initial appeal of UT. For if we need to choose a logical theory in order to even generate logical evidence, we get a kind of relativism with respect to logical evidence. To illustrate, take an arbitrary set of (formal) propositions. This set does not constitute a unique body of logical evidence, as would be natural to suppose, instead it constitutes as many different bodies of logical evidence as there are acceptable logical theories.<sup>17</sup>

This moves our discussion away from evidence—as the central topic—to a discussion of acceptable theories instead, but no such discussion should be relevant to UT. UT should not be true only relative to preferred theory. For let us remind ourselves of how strong a thesis UT really is: it concerns all bodies of evidence, no matter what subject possesses it, and no matter the time.

The crucial point is that UT is supposed to motivate a certain response to peer disagreement, i.e., at most one peer can be justified in her doxastic attitude toward the target-proposition in such disagreements. But if logical evidence is relativized to preferred logical theory, the scope of UT is reduced drastically. You can now only share logical evidence with those from your own theoretical equivalence class, and there can be as many of those classes as there are acceptable logical theories. This kind of relativism is clearly not desirable for a UT-proponent, and thus saving UT using **Objection 4** turns out to be a Pyrrhic victory.<sup>18</sup>

---

<sup>17</sup> See also Andersen (2023a) for a recent discussion of *justification holism* versus *justification atomism* in the epistemology of logic, which is highly relevant to this issue.

<sup>18</sup> Other epistemologists have suggested that one way in which uniqueness might fail is if there is a plurality of methods (in a broad sense) which one could rationally use to generate evidence. Accordingly, the counterexample **Logical Disagreement** presented here, and our discussion about logical evidence being relativized to acceptable logical theories, might be subsumed under a broader style of argument against uniqueness, namely that UT fails because evidence (of various types) is relative to acceptable methods. For further discussion of this general style of argument, see for instance Hales (2014).

Note also how the issues surrounding logical evidence and uniqueness relate to some more established debates about permissible *epistemic standards* (Titelbaum and Kopec 2019). Plenty of formal epistemologists claim that a body of evidence supports a hypothesis only relative to a rational reasoning method, and since there are multiple, extensionally non-equivalent, rational reasoning methods available, there is not always an unambiguous fact of the matter about whether some evidence supports a particular hypothesis. Subjective Bayesianism, for example, could deny UT by appealing to legitimate differences in epistemic standards. In general, Bayesians hold that any rational agent's credences at a given time can be obtained by conditionalizing their *hypothetical prior* ( $Cr_h$ ) on their total evidence at that time. For a total body of evidence  $E$  and a hypothesis  $H$ , the evidence supports the hypothesis exactly when  $Cr_h(H | E) > Cr_h(H)$ . Here, facts about evidential support are relative to the hypothetical prior of the relevant agent, and we can plausibly think of an agent's hypothetical prior as capturing their epistemic standards. Some Objective Bayesians claim that there is a unique rational

However, some might hesitate to admit that **Objection 4** leads to evidential relativism regarding logical evidence, for it may be objected that  $E$  and  $E^*$  don't have the same epistemic status. There could be good and purely epistemic reasons for favoring  $E$  over  $E^*$  (or *vice versa*) the reply goes. As noted above,  $E$  is the body of evidence induced by the tacit acceptance of classical logic, while  $E^*$  is the result of tacitly accepting intuitionist logic, but surely logicians do not just accept any old theory of logic, they have epistemic reasons for accepting whatever theory they favor. Thus,  $S_1$ 's *total* evidence pool may very well include evidence for accepting DNE, law of the excluded middle etc., which the intuitionist lacks. Similarly,  $S_2$ 's *total* evidence pool may well include evidence for denying DNE, law of the excluded middle etc., which the classical logician does not have in her possession. Further,  $S_1$ 's reasons may be better than  $S_2$ 's ditto (or *vice versa*).

Although this worry is legitimate, it will not save UT. First, it is underspecified in the literature whether UT is meant to apply to the *total* bodies of evidence in this sense, i.e., including pieces of evidence supporting one's methods used to generate evidence. There are hints about the importance of evidence for evidence-generating methods in the literature on *deep disagreement*,<sup>19</sup> but usually such evidence is taken as background information, and thus not as included in whatever body of evidence is under consideration in standard disagreement cases. Thus, it is not clear what UT-proponents would say about cases involving such *total* bodies of evidence. Further, one could easily rewrite **Logical Disagreement** stipulating that the two logicians were (known) epistemic peers. Then, insofar as evidential symmetry is necessary for peerhood, this would exclude any evidence from the case besides the common evidence. Of course, one could then say that if  $S_1$  is a classical logician and  $S_2$  an intuitionist, they cannot be epistemic peers, but in that case, we are back to square one; logical evidence becomes relativized to your own theoretical equivalence class and relativism looms.

#### 4.5 Individualistic versus Social Epistemology

**Objection 5.** UT is most plausibly defended as an *intra*-personal thesis, but **Logical Disagreement** is an inter-personal case. Thomas Kelly distinguishes between

---

hypothetical prior, so, in their case while evidential support is relative to the hypothetical prior there is still at most one rational hypothetical prior, and so UT is true. Yet some Subjective Bayesians claim that multiple hypothetical priors are rationally acceptable. Thus, for them, two rational agents could have different hypothetical priors, i.e., different epistemic standards, and end up in situations where the same body of evidence  $E$  supports a hypothesis  $H$  for one of them while it doesn't for the other.

<sup>19</sup> For detailed discussions of deep disagreement, see Lynch (2010, 2016), Kappel (2012, 2021), Ranalli (2020, 2021), Ranalli and Lagewaard (2022a, b).

*intra-personal* and *inter-personal* versions of UT (Kelly 2014, 307):

$UT_{Intra}$  Given that my evidence is  $E$ , there is some doxastic attitude  $D$  that is the only fully rational doxastic attitude for me to take towards proposition  $p$  [...].<sup>20</sup>

$UT_{Inter}$  Given evidence  $E$ , there is some doxastic attitude  $D$  that is the only fully rational doxastic attitude for anyone to take towards proposition  $p$  [...].<sup>21</sup>

Only  $UT_{Intra}$  holds as a general epistemic principle; not  $UT_{Inter}$ .

This objection saves UT as a general epistemic principle *intra-personally*, but as should be clear, it also completely undermines the core motivation for the thesis, which is social. Instead of relativizing evidence to acceptable theories or methods as in **Objection 4**,  $E$  is now relativized to subjects, and an even worse kind of relativism is unavoidable.

We should agree that  $UT_{Intra}$  is true. Take a perceptual case. If subject  $S$  clearly sees that there is a computer in front of her on the table and this visual perception constitutes her relevant evidence, then under normal circumstances there will be at most one justified doxastic attitude for her to adopt towards the proposition expressed by the sentence ‘*There is a computer on the table*,’ i.e.,  $S$  is justified in believing the proposition to be true (while either disbelieving or suspending judgement would be unjustified). Likewise,  $UT_{Intra}$  seems true in logic cases insofar as we assume the agent in play has accepted a certain logical theory (as the only correct one) in advance. This blocks cases where **Logical Disagreement** is reformulated as a single person-case with an eclectic logician who prefers neither the classical nor intuitionist tradition of logic, and yet is fully competent in both. Given our assumption, this logician cannot be *intra-personally* justified in more than one doxastic attitude towards a given  $p$ , e.g., the eclectic logician cannot be justified in a belief that  $p$  as well as a suspension of judgement with respect to  $p$  based on the same body of logical evidence.

However, as mentioned above, admitting that only  $UT_{Intra}$  is true comes with an unbearable cost for the UT-proponent. For with the embrace of this view, UT is no longer relevant to the peer disagreement debate which it was supposed to be central to. As  $UT_{Intra}$  is compatible with multiple doxastic attitudes being justified in cases of peer disagreement, the initial motivation behind UT is now completely lost. Thus, UT-proponents should not accept **Objection 5** as it indirectly undermines UT.

---

<sup>20</sup> Note that even though Kelly uses the term ‘rational’ instead of ‘justified’ in the quote above, it will not make any substantial difference for our purposes.

<sup>21</sup> See footnote 20.

## 5. Concluding Remarks

This paper has introduced a new counterexample to UT which involves logical disagreement. To legitimize this example and strengthen the case for it, we have shown that five different objections trying to save UT from **Logical Disagreement** fails. Two of the five objections were simply fended off, one needed further development to pose any real threat, while explaining away the counterexample with either one of the remaining two options resulted in an unbearable indirect defeat of the thesis. Hence, in the absence of successful objections to **Logical Disagreement**, the paper recommends that we hesitate in accepting UT as a general epistemic principle.<sup>22</sup>

## References

- Alston, W. P. (1989). *Epistemic Justification: Essays in Theory of Knowledge*. Cornell University Press.
- Andersen, F. J. (2020). “Uniqueness and Logical Disagreement”. *Logos & Episteme. An International Journal of Epistemology*, 11(1): 7–18.
- . (2023a). “Countering justification holism in the epistemology of logic: The argument from pre-theoretic universality”. *Australasian Journal of Logic*.
- . (2023b). *Logical Disagreement—An Epistemological Study*. PhD thesis, University of St Andrews.
- Beall, J. C. and Restall, G. (2000). “Logical pluralism”. *Australasian Journal of Philosophy*, 78(4): 475–493.
- . (2006). *Logical pluralism*. Oxford University Press.
- BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.
- Brouwer, L. (1975). “Intuitionism and formalism”. In *Philosophy and Foundations of Mathematics*, edited by A. Heyting, 123–138. Elsevier North-Holland.
- Carnap, R. (2014). *Logical syntax of language*. Routledge.
- Carter, J. A. and Bondy, P. (2019). *Well-founded belief: New essays on the epistemic basing relation*. Routledge.
- Conee, E. and Feldman, R. (2004). *Evidentialism: Essays in epistemology*. Clarendon Press.
- Feldman, R. (2006). “Epistemological puzzles about disagreement”. In *Epistemology Futures*, edited by S. Hetherington, 216–36. Clarendon Press.

---

<sup>22</sup> Thanks to Greg Restall, Franz Berto, Jessica Brown, Klemens Kappel, and Alessandro Rossi, for written feedback on earlier versions of the paper. My research is supported by the Niels Bohr Foundation (Royal Danish Academy of Sciences) and the Royal Institute of Philosophy (London).

- Feldman, R. and Conee, E. (1985). "Evidentialism". *Philosophical Studies*, 48(1): 15–34.
- Goldman, A. I. (1986). *Epistemology and cognition*. Harvard University Press.
- Griffiths, O. and Paseau, A. (2022). *One true logic*. Oxford University Press.
- Hales, S. D. (2014). "Motivations for relativism as a solution to disagreements". *Philosophy*, 89(1): 63–82.
- Harris, J. H. (1982). "What's so logical about the 'logical' axioms?" *Studia Logica*, 41: 159–171.
- Hattiangadi, A. (2018). "Logical disagreement". In *Metaepistemology*, edited by C. McHugh, 88–106. Oxford University Press.
- Hattiangadi, A. and Andersen, F. J. (202X). "Logical disagreement". In *The Oxford Handbook of Philosophy of Logic*. Oxford University Press.
- Hjortland, O. T. (2022). "Disagreement about logic". *Inquiry*, 65(6): 660–682.
- Kappel, K. (2012). "The problem of deep disagreement". *Discipline Filosofiche*, 22(2): 7–25.
- . (2021). "Higher order evidence and deep disagreement". *Topoi*, 40(5): 1039–1050.
- Kauss, D. (2023). "A rational agent with our evidence". *Erkenntnis*.
- Kelly, T. (2010). "Peer disagreement and higher order evidence". In *Disagreement*, edited by R. Feldman, and T.A. Warfield, 111–175, Oxford University Press.
- . (2014). "Evidence can be permissive". In *Contemporary debates in epistemology*, edited by M. Steup, J. Turri, and E. Sosa, 298–311. Wiley Blackwell.
- Kopec, M. and Titelbaum, M. G. (2016). "The uniqueness thesis". *Philosophy Compass*, 11(4): 189–200.
- Korcz, K. A. (2021). "The Epistemic Basing Relation". In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University, Spring 2021 edition.
- Littlejohn, C. (2012). *Justification and the truth-connection*. Cambridge University Press.
- Lynch, M. (2010). "Epistemic circularity and epistemic disagreement". In *Social epistemology*, edited by A. Haddock, P. D. Millar, and D. Pritchard, 262–277. Oxford University Press.
- . (2016). "After the spade turns: disagreement, first principles and epistemic contractarianism". *International Journal for the Study of Skepticism*, 6(2-3): 248–259.
- Matheson, J. (2011). "The case for rational uniqueness". *Logos & Episteme. An International Journal of Epistemology*, 2(3): 359–373.

Frederik J. Andersen

- McCain, K. (2014). *Evidentialism and epistemic justification*. Routledge.
- Quine, W. V. O. (1986). *Philosophy of logic*. Harvard University Press.
- Ranalli, C. (2020). “Deep disagreement and hinge epistemology”. *Synthese*, 197(11): 4975–5007.
- . (2021). “What is deep disagreement?” *Topoi*.
- Ranalli, C. and Lagewaard, T. (2022a). “Deep disagreement (part 1): Theories of deep disagreement”. *Philosophy Compass*, 17(12): e12886.
- . (2022b). “Deep disagreement (part 2): Epistemology of deep disagreement”. *Philosophy Compass*, 17(12): e12887.
- Rosa, L. (2012). “Justification and the uniqueness thesis”. *Logos & Episteme. An International Journal of Epistemology*, 3(4):571–577.
- . (2016). “Justification and the uniqueness thesis again: A response to Anantharaman”. *Logos & Episteme. An International Journal of Epistemology*, 7(1):95–100.
- Ross, R. (2021). “Alleged counterexamples to uniqueness”. *Logos & Episteme. An International Journal of Epistemology*, 12(2): 203–213.
- Rossi, A. (2023). “From collapse theorems to proof-theoretic arguments”. *Australasian Journal of Logic*, 20(1): 1–31.
- Russell, G. (2018). “Logical nihilism: Could there be no logic?” *Philosophical Issues*, 28(1): 308–324.
- . (2019). “Logical pluralism”. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Metaphysics Research Lab, Stanford University, summer 2019 edition.
- Shapiro, S. (2014). *Varieties of logic*. Oxford University Press.
- Silva, P. (2023). “Merely statistical evidence: when and why it justifies belief”. *Philosophical Studies* 180 (9): 2639–2664.
- Smart, J. A. (2021). “Disbelief is a distinct doxastic attitude”. *Synthese*, 198(12): 11797–11813.
- Titelbaum, M. (2015). “Rationality fixed point (or: In defense of right reason)”. In *Oxford Studies in Epistemology* 5, edited by T. S. Gendler, and J. Hawthorne, 253–294. Oxford University Press.
- . (2019). “Return to reason”. In *Higher-order evidence. New essays*, edited by M. Skipper, and A. Steglich-Petersen, 226–245. Oxford University Press.
- Titelbaum, M. G. and Kopec, M. (2019). “When rational reasoners reason differently”. In *Reasoning: New Essays on Theoretical and Practical Thinking*, edited by M. B. Jackson, B. B. Jackson, 205–231. Oxford University Press.

- White, R. (2014). "Evidence cannot be permissive". In *Contemporary debates in epistemology*, edited by M. Steup, J. Turri, and E. Sosa, 312–323. Wiley Blackwell.
- Williamson, T. (1988). "Equivocation and existence". In *Proceedings of the Aristotelian Society* 88: 109–127.
- . (2000). *Knowledge and its Limits*. Oxford University Press.
- . (2007). *The philosophy of philosophy*. Wiley Blackwell.





# IS JTB KNOWLEDGE HOPELESS?

Arnold CUSMARIU

**ABSTRACT:** An argument structure that covers both cases Gettier described in his 1963 paper reinforces the conclusion of my 2012 *Logos & Episteme* article that the justified true belief (JTB) conception of knowledge is inconsistent. The stronger argument makes possible identification of fundamental flaws in the standard approach of adding a fourth condition to JTB, so that a new kind of skepticism becomes inevitable unless conceptual change occurs.

**KEYWORDS:** Gettier Problem, inconsistency, JTB knowledge, ostensible evidence, skepticism

## Background

In articles I published in *Logos & Episteme* in 2012 and 2016, I presented and defended a semantic approach in epistemology intended to capture arguably our most valuable repositories of knowledge, mathematics and science. The 2012 article also stated one of two components of the Gettier Problem (GP) in argument form and explained how semantic epistemology dealt with GP. I am returning to the issue to formulate a stronger argument that covers both GP components, which reinforces the earlier conclusion that the justified true belief (JTB) conception of knowledge is inconsistent and reveals a fundamental flaw in the standard approach to GP of adding a fourth condition to JTB. A new kind of skepticism becomes inevitable unless conceptual change occurs.

## A Precedent Missed

The literature on the Gettier Problem (GP) has been growing steadily since the 1963 publication of Edmund L. Gettier's paper. Books, articles and collections of article have analyzed the problem and proposed solutions. A lengthy and detailed survey by Stephen Hetherington in the *Internet Encyclopedia of Philosophy*<sup>1</sup> assessed contributions and remaining challenges.

Missing from the GP literature is awareness of a remarkably similar event some sixty years earlier that had even more devastating consequences. I refer to Bertrand Russell's proof that Rule V in Frege's *Grundgesetze der Arithmetik* was

---

<sup>1</sup> <https://iep.utm.edu/gettier/#H14/>.

inconsistent. In a 1902 letter to Frege (van Heijenoort 1967, 124-5), Russell showed that substituting the open sentence " $\sim(x \in x)$ " in Rule V entails contradiction.<sup>2</sup> Two important lessons were learned from Russell's Paradox, both missed by the GP literature. First, counterexamples are inconsistency proofs and as such should be spelled out in argument form. Second, conceptual change may prove necessary.<sup>3</sup> I will address both issues.<sup>4</sup>

### Gettier Text Analysis

Gettier presents two counterexamples<sup>5</sup> to the analysis of knowledge as justified true belief (JTB). It is not explained why two counterexamples are necessary—why not three?—nor the difference between them, the implication being that two counterexamples are needed so that different objections can be raised. This is false as we shall see.

Gettier states at the outset that he will "argue"—as opposed to merely state, contend, hypothesize or claim—that a certain proposition is false, namely, that justified true belief is "*sufficient*" (1963, 121, original italics) for knowledge. As he notes, this proposition is one conjunct of JTB, according to which justified true belief is necessary and sufficient for knowledge. Gettier uses the term "argument" in the very next sentence, implying intent to defend the proposition that justified true belief is insufficient for knowledge by means of standard logic. The paper states at the end (1963, 123) that "the two examples show" that justified true belief is not sufficient for knowledge, implying by the use of "show" that an argument has been presented for this conclusion.

So, do we find in Gettier's paper an argument in the standard logical sense whose conclusion is that JTB is insufficient for knowledge?

We do not.

---

<sup>2</sup> Inconsistent sentences can be of the form " $p \& \sim p$ " or of the form " $p \equiv \sim p$ ". Russell proved the latter and went on to solve the problem he had discovered. Gettier never "returned to the scene of the crime."

<sup>3</sup> The naive conception of a set was eventually replaced with the iterative conception of a set (see Boolos 1971). I attempted to rescue naive sets using a "patch" approach in Cusmariu 1979.

<sup>4</sup> On a personal note, I recall an episode in a Chisholm seminar at Brown University in the mid-70s during which he recalled his first reaction to GP, frustration evident in his tone: "Well, all you have to do is say this ..." Chisholm never stated GP in argument form, in class or in his books; nor did he seem to realize that a much more devastating conclusion had been proved. Like everyone else at the time, he focused on repairing the analysis.

<sup>5</sup> Gettier does not use the term "counterexample"; he uses the term "Cases." The earliest use of the term "counterexample" seems to be in Clark 1963.

What we have instead are what are called “points,” which Gettier says he is “noting.” Here is the language (1963, 121):

I shall begin by noting two points. First, in the sense of “justified” in which S’s being justified in believing P is a necessary condition of S’s knowing that P, it is possible for a person to be justified in believing a proposition that is in fact false. Secondly, for any proposition P, if S is justified in believing P, and P entails Q, and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q.

This passage is a muddle. For example, what are we to make of the locution “in the sense of ‘justified’ in which ...” What sense is that? The text does not say.

To get a handle on what is being asserted, let us list component propositions:

- (a) If S knows P, then S is justified in believing P.
- (b) If S is justified in believing P, and P entails Q, and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q.
- (c) It is possible for a person to be justified in believing a proposition that is false.

The passage explicitly links (c) with (a), leaving (c)-(b) links up in the air. I can think of three ways (c) and (a) could be linked, all of which seem to me false:

- 1. Propositions (a) is meaningful only if proposition (c) is true.
- 2. Propositions (a) is true only if proposition (c) is true.
- 3. Propositions (a) entails proposition (c).

If not 1-3, how are (a) and (c) to be linked? The text does not say.

What about (c) and (b)? Should (b) likewise be read this way:

- (b\*) If S is justified in believing P, and P entails Q, and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q “in the sense of ‘justified’ in which it is possible for a person to be justified in believing a proposition that is false”?

The text offers no comment. Interpretations 1-3 are also false if applied to (b).

I propose ignore these complications and interpret the first “point” of the passage as stating as an assumption that it is possible for a proposition to be justified and false, omitting potential links to (a) and (b), which are unnecessary for the purpose of building an argument. As to what “possible” means, we can make do with “we can imagine circumstances in which ...” the circumstances being Gettier’s two Cases.

Arnold Cusmariu

## A Two-Stage Argument

The same argument structure is sufficient to cover both Cases.

### Stage I: Proving that Smith knows that Q.

#### 1. The proposition that P is evident for Smith.

Case I: P is the proposition that Jones will get the job, and Jones has ten coins in his pocket.

*Comment:* Gettier states (p. 122) that P is evident for Smith because “the president of the company assured him that Jones would in the end be selected” and because Smith “counted the coins in Jones's pocket ten minutes ago.”

Case II: P is the proposition that *Jones owns a Ford.*

*Comment:* Gettier states (1963, 122) that P is evident for Smith because “Jones has at all times in the past within Smith's memory owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford.”

#### 2. P logically implies Q.

Case I: Q is the proposition that the man who will get the job has ten coins in his pocket.

*Comments:* Gettier describes (1963, 122) the logical connection between P and Q as entailment, suggesting that P logically implies Q. This is not obvious.

- Proposition P has the form  $((\exists x)(\exists y)((x=a \ \& \ Fx) \ \& \ (y=b \ \& \ Gy) \ \& \ x=y))$  – where  $a$  and  $b$  are individual constants co-designating Smith and  $F$  and  $G$  are the predicates “will get the job” and “has ten coins in his pocket,” respectively.
- Proposition Q has the form  $((\exists!x)(\exists!y)(Fx \ \& \ Gx) \ \& \ (Fy \ \& \ Hy) \ \& \ x=y)$  – where  $\exists!$  is a uniqueness quantifier,  $F$  and  $G$  are the predicates “is a man” and “will get the job,” respectively; and  $H$  is the predicate “has ten coins in his pocket.”
- It seems clear that logical form by itself is not sufficient to show that P logically implies Q. However, let’s not worry about additional assumptions needed to get the entailment to go through and grant that the doxastic burden is not so onerous that Smith cannot recognize the entailment.

Case II: Q is the proposition that Jones owns a Ford or Brown is in Barcelona.

*Comment 1:* Gettier mentions two more versions Q, about Brown being in Boston or in Brest-Litovsk. However, Gettier focuses only on the Barcelona Q, so I propose to ignore the other versions.

*Comment 2:* Q is follows logically from P as a substitution instance of an elementary rule of the propositional calculus, *Disjunction*.

3. Smith recognizes the inference from P to Q.

4. Smith accepts Q.

*Comment:* Gettier states that Smith accepts Q in both Cases on the basis of recognizing the inference from P to Q.

5. Q is true.

6. For any propositions X and Y, if X is evident for a person S, X logically implies Y, S recognizes the inferences from X to Y, and S accepts Y on the basis of recognizing this inference, then Y is evident for S.

*Comment:* Step 6 is Gettier second “point.”

Therefore, by instantiation and *Modus Ponens* from Steps 1, 2, 3, 4 and 6:

7. Q is evident for Smith.

8. The JTB conception of knowledge entails the following:

- For any proposition X, if X is true, S accepts X, and X is evident for S, then S knows X.

Therefore, by instantiation and *Modus Ponens* from Steps 5, 4, 7 and 8:

9. Smith knows that Q.

### Stage II: Proving that Smith does not know that Q.

We start by citing the language Gettier provides to support his claim that Smith does not know that Q in both Cases.

#### Case I (1963, 122):

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not know that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

Let us use our P and Q symbolism and link to Gettier's.

- P is Gettier's (d), the proposition that Jones will get the job, and Jones has ten coins in his pocket.

*Comment:* Recall that P has this logical form:  $((\exists x)(\exists y)((x=a \ \& \ Fx) \ \& \ (y=b \ \& \ Gy) \ \& \ x=y))$  – where *a* and *b* are individual constants co-designating Smith and *F* and *G* are the predicates “will get the job” and “has ten coins in his pocket,”

Arnold Cusmariu

respectively.

- Q is Gettier's (e), the proposition that the man who will get the job has ten coins in his pocket.

*Comment:* Recall that Q has this logical form:  $((\exists!x)(\exists!y)(Fx \& Gx) \& (Fy \& Hy) \& x=y)$  – where  $\exists!$  is the uniqueness quantifier,  $F$  and  $G$  are the predicates “is a man” and “will get the job,” respectively; and  $H$  is the predicate “has ten coins in his pocket.”

Using our P and Q symbolism yields this paraphrase:

But it is equally clear that Smith does not know that Q is true; for Q is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in Q on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

*Comment:* I will simplify this convoluted explanation shortly.

Case II (1963, 123):

But imagine now that two further conditions hold. First, Jones does not own a Ford, but is at present driving a rented car. And secondly, by the sheerest coincidence, and entirely unknown to Smith, the place mentioned in proposition (h) happens really to be the place where Brown is. If these two conditions hold then Smith does not know that (h) is true, even though (i) (h) is true, (ii) Smith does believe that (h) is true, and (iii) Smith is justified in believing that (h) is true.

Let us use our P and Q symbolism here as well and link to Gettier's.

- P is Gettier's (f), the proposition that *Jones owns a Ford*.
- Q is Gettier's (h), the proposition that *Jones owns a Ford or Brown is in Barcelona*.

Using our P and Q symbolism yields this paraphrase:

But imagine now that two further conditions hold. First, P is false. And secondly, by the sheerest coincidence, and entirely unknown to Smith, the place mentioned in proposition Q happens really to be the place where Brown is. If these two conditions hold then Smith does not know that Q is true.

We are ready to start the argument at Stage II.

11. If Q is true by coincidence (accident, happenstance, luck) and Smith bases his belief that Q on an inference from a false but evident belief that P, then Smith does not know that Q.

*Comment:* Step 11 sums up the factors that, on my interpretation, drive Gettier's claim that Smith does not know Q in both Cases.

12. Q is true by coincidence (accident, happenstance, luck) in both cases.

- Case I: Gettier has established that it is a coincidence that Smith happens to have ten coins in his pocket.
- Case II: Gettier has established that it is a coincidence that Brown happens to be in Barcelona.

13. Gettier has established that Smith based Q on an inference from P in both Cases.

14. P is false but evident in both Cases.

- Case I: Gettier has established that it is evident for Smith that Jones will get the job, adding that Jones will not get the job; so that in the circumstances described, it is false but evident that Jones will get the job.
- Case II: Similarly, Gettier has established that it is evident for Smith that Jones owns a Ford, adding that Jones no longer owns a Ford but is driving a rented car, meaning that in the circumstances described; so that it is false but evident that Jones owns a Ford.

Therefore, by instantiation and *Modus Ponens* from Steps 11-14:

15. Smith does not know that Q in either Case.

Therefore, by *Conjunction*:

16. In both Cases, Smith knows that Q and Smith does not know that Q.

17. The JTB conception of knowledge is inconsistent.

## Ways Out

### 1. So What?

In epistemology and analytic philosophy generally, studying a concept means providing an analysis in terms of necessary and sufficient conditions—NS for short. The following considerations apply.

- I. An NS is possible for any concept.
- II. An NS is necessary for the purpose of applying a concept correctly.
- III. A person must be in possession of an NS to apply a concept correctly.
- IV. Something of practical value is gained by having an NS of a concept.
- V. Something of theoretical value is gained by having an NS of a concept.

I-IV are false or doubtful. Responding to GP by rejecting V, i.e., doing without an NS of propositional knowledge, seems extreme or at least premature.

Arnold Cusmariu

For one thing, it would not answer Plato's question in the *Meno* about the difference between knowledge and true belief.

## 2. The "Patch" Approach

The standard approach has been to accept that the JTB conception of knowledge has been refuted and attempt repairs by adding a fourth condition.<sup>6</sup> However, the next Gettier counterexample may be just around the corner—the "wild goose chase" worry—leading to increasingly complicated JTB+ conceptions of knowledge, so complicated that they risk distorting pre-analytic intuitions to the point that one cannot tell what knowledge means under various revisions or how to test them.

To put matters another way, consider the contrapositive of 11:

11a. If Smith knows that Q, then either it is not the case that Q is true by coincidence (accident, happenstance, luck) or it is not the case that Smith bases his belief that Q on an inference from a false but evident belief that P.

A repair of JTB that implemented the consequents of Step 11—of which there are quite a few in the literature (see Hetherington)—is not a substitute for a proof of consistency. The GP literature is not even aware that such a proof is necessary.

## 3. The Ostensibly Evident

Regarding Gettier's first "point," I'd like to suggest that we restrict "evident" to true propositions and apply "ostensibly evident" to false propositions. This has the following consequences:

- Case I: The proposition that Jones is the man who will get the job is only ostensibly evident for Smith if Smith will get the job rather than Jones.
- Case II: The proposition that Jones owns a Ford is only ostensibly evident for Smith if Jones does not own a Ford but is driving a rented car.
- An epistemic principle whose antecedent contains ostensibly evident propositions and can only transmit ostensible evidence. Thus, if P is false in both Cases, Step 6 (Gettier's second "point") must be restated to read "For any propositions X and Y, if X is ostensibly evident for a person S, X logically implies Y, S recognizes the inference from X to Y, and S accepts Y on the basis of recognizing this inference, then Y is ostensibly evident for S."

---

<sup>6</sup> That is the approach taken by Chisholm, one of Gettier's targets, in the three editions of his *Theory of Knowledge* (1966, 23; 1977, 105-110; and 1989, 91-100). Interestingly, A.J. Ayer, the other philosopher Gettier targeted, seems to have shown no interest in answering GP; at least, not in print.



- Accordingly, Step 8 must be restated to read “For any proposition X, if X is true, S accepts X, and X is ostensibly evident for S, then S ostensibly knows X.”
- So, GP is about a new category of ostensible knowledge, the other category being claims to know something that turns out to be false.
- False evidence also raises this worry: While Q is true in both Gettier Cases, these propositions have been inferred from false but evident propositions so, in a sense, they are... vacuously true!
- Look at it this way: We cannot be said to know what is false; so how is it that we can be said to know on the basis of false propositions?

### Wider Implications

Some of the consequences of restricting “evident” to truths in a semantic conception of scientific and mathematical knowledge are addressed in Cusmariu 2012 and 2016. Exploring the implications of this restriction epistemology-wide would require book-length treatment. Two other issues are topics for another time:

- How to prove that the semantic conception of knowledge is consistent; and, if it is consistent, whether it is complete in mathematics in light of Gödel’s discoveries.
- Whether epistemic properties are properties of propositions according to semantic epistemology or properties of persons according to virtue epistemology (see Baehr, Code 1987, Sosa 1991, Plantinga 1992, and Montmarquet 1993).

The concept of ostensible evidence deserves consideration independently of the Gettier Problem—just as iterative sets would have been a good idea independently of the Russell Paradox.<sup>7</sup> Should it turn out, however, that epistemology is stuck with a “patch” approach, an argument for skepticism would result to the effect that the JTB knowledge has gaps of unknowable number—Gettier Cases—and therefore we JTB-know less than we think we know. This kind of skepticism is very different from the kind philosophers have considered since Descartes and might well prove to be the proverbial “straw that breaks the camel’s back” of the JTB conception of knowledge.<sup>8</sup>

---

<sup>7</sup> I would argue that any theory proposed as a solution to a problem should satisfy this requirement, in science as well as philosophy.

<sup>8</sup> David Christensen, Christopher Hill and Gary Rosenkrantz provided helpful comments on earlier drafts.

## References

- Baehr, Jason S. No date. "Virtue Epistemology." Internet Encyclopedia of Philosophy <https://iep.utm.edu/virtue-epistemology/>.
- Code, Lorraine. 1987. *Epistemic Responsibility*. Hanover, NH: University Press of New England.
- Cusmariu, Arnold. 2012. "Toward a Semantic Approach in Epistemology." *Logos & Episteme. An International Journal of Epistemology* III(4): 531-543.
- . 2016. "Semantic Epistemology Redux: Proof and Validity in Quantum Mechanics." *Logos & Episteme. An International Journal of Epistemology* VII(3): 287-304.
- . 1979. "Russell's Paradox Re-examined." *Erkenntnis* 14: 365-370.
- Boolos, George. 1971. "The iterative conception of set." *The Journal of Philosophy*, 68(8): 215-231.
- Borges, Rodrigo, De Almeida, Claudio & Klein, Peter D. (eds). 2017. *Explaining Knowledge: New Essays on the Gettier Problem*. Oxford: Oxford University Press.
- Chisholm, Roderick M. 1966. *Theory of Knowledge*. Englewood Cliffs, NJ, Prentice Hall.
- . 1977. *Theory of Knowledge*. Englewood Cliffs, NJ, Prentice Hall.
- . 1989. *Theory of Knowledge*. Englewood Cliffs, NJ, Prentice Hall.
- Clark, Michael. 1963. "Knowledge and Grounds: A Comment on Mr. Gettier's Paper," *Analysis* 24: 46-8.
- Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121-123.
- Hetherington, Stephen. No date. "Gettier Problems." Internet Encyclopedia of Philosophy <https://iep.utm.edu/gettier/#H14>.
- Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.
- Roth, Michael D & Galis, Leon (eds). 1984. *Knowing: Essays in the Analysis of Knowledge*. Lanham MD: University Press of America.
- Shope, Robert K. 2017. *The Analysis of Knowing: A Decade of Research*. Princeton: Princeton University Press.
- Sosa, Ernest. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press.
- van Heijenoort, Jean (ed.) 1967. *From Frege to Gödel: A Sourcebook in Mathematical Logic, 1879-1931*. Cambridge: Harvard University Press.

# WHEN THE (BAYESIAN) IDEAL IS NOT IDEAL

Danilo Fraga DANTAS

**ABSTRACT:** Bayesian epistemologists support the norms of probabilism and conditionalization using Dutch book and accuracy arguments. These arguments assume that rationality requires agents to maximize practical or epistemic value in every doxastic state, which is evaluated from a subjective point of view (e.g., the agent's expectancy of value). The accuracy arguments also presuppose that agents are opinionated. The goal of this paper is to discuss the assumptions of these arguments, including the measure of epistemic value. I have designed AI agents based on the Bayesian model and a nonmonotonic framework and tested how they achieve practical and epistemic value in conditions in which an alternative set of assumptions holds. In one of the tested conditions, the nonmonotonic agent, which is not opinionated and fulfills neither probabilism nor conditionalization, outperforms the Bayesian in the measure of epistemic value that I argue for in the paper ( $\alpha$ -value). I discuss the consequences of these results for the epistemology of rationality.

**KEYWORDS:** bounded rationality, computational epistemology, Bayesian epistemology, epistemic utility theory, nonmonotonic logic

## Introduction

Bayesian epistemologists propose the norms of probabilism and conditionalization.<sup>1</sup> The initial justifications for these norms were Dutch book arguments (DBAs). A DB is a series of bets, each of which the agent considers fair, but when taken together, they result in a guaranteed loss. A DBA aims to demonstrate that if an agent violates a specific Bayesian norm, then it is possible to create a DB against her. Ramsey (1926) and De Finetti (1937) proposed DBAs for probabilism, and Teller (1973) for conditionalization. Satisfaction of a given Bayesian norm does not guarantee invulnerability to DBs for other norms, but if a reasoner fulfills a norm (e.g., probabilism), then it can be shown that a DB for that norm cannot be constructed against her (e.g., Kemeny 1955). Since vulnerability to DBs may lead an agent to buy and sell a series of bets that amounts to a sure loss, it is associated

---

<sup>1</sup> Probabilism states that a rational agent's credence function is consistent with the axioms of probability. Conditionalization states that a rational agent who becomes certain of some new piece of evidence updates her previous unconditional credences by conditionalizing them on the evidence (see Talbott 2016).

with the minimization of ‘practical value,’ which, in turn, is related to the fulfillment of the agent’s practical goals (e.g., having monetary gain). Some critics argue that vulnerability to DBs is not adequately related to the minimization of practical value.<sup>2</sup> Others argue that *practical* rationality might be associated with maximizing practical value, while *epistemic* rationality should be associated with maximizing ‘epistemic value.’

Different features of a set of ‘beliefs’<sup>3</sup> are putative sources of epistemic value: closure, coherence, amount of evidential support, etc. In the last decades, a group of epistemologists has argued for ‘veritism,’ the thesis that the fundamental source of epistemic value is the believing of truths rather than falsehoods.<sup>4</sup> These epistemologists sought to justify the Bayesian norms in purely epistemic terms (i.e., veritistic), which resulted in the epistemic utility theory (EUT). The ‘accuracy arguments’ of EUT purport to show that the fulfillment of the Bayesian norms leads to the minimization of (expected) epistemic disvalue (inaccuracy, see sec. 1). Joyce (1998) argues that if an agent’s credence function is not probabilistic, then she is in the position to know that there is a probabilistic function that is less inaccurate than hers in all situations that she finds possible. Leitgeb and Pettigrew (2010) argue that an agent who updates her credences using a non-Bayesian strategy is in the position to know that an update using conditionalization minimizes expected inaccuracy relatively to hers, where this expectancy is calculated using her credences.

I use the expression ‘maximization of value’ loosely, in such a way that it includes the avoidance of minimization of value (e.g., the avoidance of a ‘sure loss’) and the minimization of (expected) disvalue. In this sense, the DBAs and the accuracy arguments are ‘maximization arguments,’ as they evaluate an agent’s rationality in terms of how much they maximize practical or epistemic values. These arguments assume the following:

- a1. Rationality requires maximization of value in every doxastic state;
- a2. ‘Maximization of value’ is evaluated from a subjective point of view;
- a3. Agents are opinionated.

---

<sup>2</sup> For example, Hájek (2009, 231) points out that violating the Bayesian norms amounts not only to the construction of DBs but also to Czech books (a series of fair bets that together amount to a sure win).

<sup>3</sup> I use ‘belief’ as a general term, which encompasses both full beliefs and credences.

<sup>4</sup> For example: “the fundamental source of epistemic value for a doxastic state is the extent to which it represents the world correctly: that is, its fundamental epistemic value is determined entirely by its truth or falsity” (Pettigrew 2019, 761). See Littlejohn (2015) for a critical discussion about veritism.

These arguments assume a1 because they concern the doxastic states of an agent at an arbitrary point in time (e.g., after an update). Failing to maximize value at that time is sufficient for irrationality. Nevertheless, a rational agent might risk value at the present if this may result in a gain of value in the future. A related issue is that the DBAs assume that a rational agent cannot make choices that lead to a sure loss of any practical value (monetary gain is not a priority practical value). Nevertheless, a rational agent may be forced to make such choices when facing trade-offs between goals. These arguments assume a2 because they measure value in such a way that is accessible to the agents. For example, Leitgeb and Pettigrew measure the expectation of value using the agents' credences.<sup>5</sup> This assumption is reasonable because rationality is usually assumed to have a subjective character (e.g., Wedgwood 2015, 221), but whether agents maximize actual value should also interest epistemology.<sup>6</sup> Opinionation is the holding of belief-values for every proposition in the agent's agenda (the set of propositions of interest). The accuracy arguments assume a3 because they always compare sets of beliefs with values for the same propositions (the agenda). This assumption is related to the difficulty of using inaccuracy to compare sets of beliefs with different sizes (see sec. 1). Nevertheless, it may be rational to 'withhold beliefs'<sup>7</sup> in some situations (e.g., in the absence of evidence).

I intend to investigate how much of the justificatory force of the maximization arguments depends on a1-a3 and whether these are necessary assumptions in a study about rationality. The investigation will proceed by comparing how a Bayesian agent and an agent who does not fulfill the Bayesian norms maximize practical and epistemic values in conditions in which a1-a3 do not hold. Several formal systems model uncertain reasoning in a way that deviates from the tenets of probability (see Genin and Huber 2020). The nonmonotonic frameworks are among those that deviate the most. There is a myriad of nonmonotonic frameworks, but not all of them are far enough from probability

---

<sup>5</sup> The DBAs appeal to the notion of sure loss and Joyce's argument appeals to the principle of dominance. These elements are objective in the sense of concerning guaranteed outcomes. But these arguments still conform to a2 because the agents are in the position to know that those outcomes are guaranteed.

<sup>6</sup> "[I]t would seem absurd to claim that it is epistemically more important to have an update rule that minimizes *expected* inaccuracy than to have one that *actually* minimizes inaccuracy" (Douven 2013, 438, his emphases).

<sup>7</sup> I use 'withhold beliefs' to denote the attitude of neither believing nor disbelieving a proposition. I am not using this expression to denote the attitude of holding middling credences about a proposition (see Sturgeon 2008).

(for our purposes).<sup>8</sup> I will focus on a three-valued version of logic programming (LP, see Doets 1994), whose third value may be used for modeling non-opinionated agents (against a3). The comparison will be performed using the tools of computational epistemology (e.g., Douven 2013), where epistemologists design computer simulations of AI agents interacting with environments that are randomly generated from fixed parameters. The measures of practical and epistemic values will be done only at the end of each trial (against a1). ‘Value in a random trial’ (actual value) is a contingent notion, but, if the number of trials is large enough, their mean approximates an ‘objective expected value’ (against a2), which results from actual values and environmental probabilities and not the agent’s beliefs.

I have designed computer simulations of two AI agents facing an epistemic version of the Wumpus World, a class of environments used for investigating uncertain reasoning. The first is the ‘Bayesian agent,’ who holds degrees of belief (credences), is opinionated, and fulfills both probabilism and conditionalization. The second is the ‘nonmonotonic agent,’ based on a three-valued LP, who holds all-or-nothing beliefs (full beliefs), is not necessarily opinionated, and fulfills neither probabilism nor conditionalization. I analyze how often these agents solve the problem (practical rationality) and the epistemic value of their beliefs (epistemic rationality).<sup>9</sup> The idea is not to refute the conclusions of Bayesian epistemology, but to discuss the assumptions in the notion of maximization of (practical or epistemic) value, including the measure of epistemic value. In Section 1, I introduce the Wumpus World and the measures of epistemic value ( $\alpha$ - and  $\beta$ -values, with linear, Brier, and log scoring rules) and of practical value (p-value) used in the investigation. In Section 2, I discuss the implementation of these models in a computer simulation and present the results from the simulation. In one condition (c3), the nonmonotonic agent outperforms the Bayesian in  $\alpha$ -value. In Section 3, I discuss the measures of epistemic value, whether a1-a3 are necessary assumptions in a study about rationality, and how the results from EWW impact the epistemology of rationality.

---

<sup>8</sup> For example, Bourne and Parsons (1998) show how to use System P (Kraus et al. 1990) to propagate the lower bounds on conditional probabilities.

<sup>9</sup> The study of practical rationality most often concerns the principles of choice between actions given the beliefs and preferences of an agent (e.g., in decision theory, see Steele and Stefánsson 2020). I use ‘practical rationality’ in the broader sense of ‘how much the agent’s beliefs assist her practical goals.’

## 1. Setting the Stage

The Wumpus World (WW, see Russell and Norvig 2020, 210) is a class of environments used in AI for studying uncertain reasoning. WW is a cave comprising rooms connected by passageways and surrounded by walls. Somewhere in the cave, there is the Wumpus, a beast that kills anyone who enters its room. Other rooms contain bottomless pits that kill anyone who steps in. Finally, there is a heap of gold hidden somewhere in the cave. The agent's goal is to explore the cave, find the gold, and escape the cave with the gold. The agent cannot enter a room containing a pit or the Wumpus without dying, but the rooms around those contain a breeze and a stench (respectively). WW is an interesting environment for studying uncertain reasoning because it is only partially observable, in such a way the absence of direct information about pits and the Wumpus forces the agents to draw provisional conclusions and revisit them given new information. WW is more complex than the average environment used in computational epistemology, but it also is more 'realistic' than the average because it associates a practical cost (risk of death) with the gathering of evidence. This feature enables us to integrate the investigation of practical and epistemic rationality.

The only goal of WW ('grab the gold and escape the cave') is a practical goal, but I am also concerned with epistemic rationality. For this reason, I have developed an epistemic version of WW (EWW). EWW has the additional goal of forming the most comprehensive and accurate set of beliefs about the positions of the pits and the Wumpus. EWW has an additional layer of uncertainty: the agent may perceive 'random' breezes and stenches, which are independent of the pits and the Wumpus. The occurrence of random breezes and stenches may be interpreted as a feature of the agent's perception, which may return a persistent false positive (a perceptual illusion).<sup>10</sup> Random breezes and stenches may be used to evaluate how reasoning processes cope with a faulty perception. EWW may be represented as a matrix of dimensions  $s \times s$ , where the room  $(x, y)$  is in the  $x$ th row and  $y$ th col of the matrix. In the following,  $p_{x,y}$  states that there is a pit in  $(x, y)$ . Similarly,  $w_{x,y}$  for Wumpus,  $g_{x,y}$  for gold,  $b_{x,y}$  for breezes, and  $s_{x,y}$  for stenches. The Wumpus and the gold are randomly placed in a room other than  $(0, 0)$ . Rooms

---

<sup>10</sup> The most natural model of a faulty perception is one in which each perception has some probability of returning a non-persistent false positive/negative (e.g., about breezes). This model introduces some complications for our purposes. For example, a Bayesian agent with faulty perception in this sense would need to update her credences using Jeffrey conditionalization (Jeffrey, 1983), but Leitgeb and Pettigrew (2010) have shown that Jeffrey conditionalization does not necessarily minimize expected inaccuracy. This result is exploited by Trpin and Pellert (2019), who use the natural model of faulty perception.

other than (0, 0) contain a pit with probability  $pr(p)$  and a random breeze or random stench with an independent probability of  $pr(rand)$ .<sup>11</sup> The agent starts in (0, 0) (‘the entrance’), facing east. She knows the description of EWW and the size of the cave, but she ignores the configuration of the cave (the position of pits, Wumpus, and gold).

The most distinguishing feature of EWW is the measure of epistemic value. The measure used in EUT is one of epistemic disvalue (inaccuracy), defined as the ‘distance’ between a set of beliefs and the ‘ideal’ set containing beliefs about the same propositions (see Carr 2015, 223),<sup>12</sup> where the belief that  $\varphi$  has the value 1 when  $\varphi$  is true and 0 when it is false. Inaccuracy is not an adequate measure of epistemic disvalue for the comparison between the Bayesian and the nonmonotonic agents because the latter can withhold beliefs, which is a ‘cheap’ way of minimizing inaccuracy. If inaccuracy was the measure of epistemic disvalue, then the nonmonotonic agent would have no reason to explore the cave and form new beliefs because, in doing so, she would risk increasing (but not decreasing) her inaccuracy (Dantas 2022). The comparison between those agents requires a measure of epistemic value that awards comprehensiveness and accuracy, which may be done by measuring the amount of truth ( $t$ ) and of falsehood ( $f$ ) in the agent’s belief-set and then conjoining these values using an ‘alethic’ function  $a(t, f)$ .

	$t_\varphi$	$f_\varphi$
Linear score	$1 - \varepsilon_\varphi$	$\varepsilon_\varphi$
Brier score	$(1 - \varepsilon_\varphi)^2$	$(\varepsilon_\varphi)^2$
Logarithmic score	$-\ln(\varepsilon_\varphi)$	$-\ln(1 - \varepsilon_\varphi)$

Table 1: Scoring rules, where  $\varepsilon_\varphi = |\nu(\varphi) - b(\varphi)|$ ,  $\nu(\varphi)$  is  $\varphi$ ’s truth-value, and  $b(\varphi)$  is the belief-value of  $\varphi$  for the agent.

The values of  $t$  and  $f$  may be measured using different ‘scoring rules.’ These values are the global amounts of truth and falsehood in the agent’s set of beliefs, measured as the sum of the local amounts  $t_\varphi$  and  $f_\varphi$ , for every proposition  $\varphi$  in the agent’s belief-set  $B$  (i.e.,  $t = \sum_{\varphi \in B} t_\varphi$  and  $f = \sum_{\varphi \in B} f_\varphi$ ). Let the ‘error’ of the belief that  $\varphi$  be  $\varepsilon_\varphi = |\nu(\varphi) - b(\varphi)|$ , where  $\nu(\varphi)$  is  $\varphi$ ’s truth-value and  $b(\varphi)$  is  $\varphi$ ’s belief-value for

<sup>11</sup> I have also introduced some simplifications to the traditional WW. No room contains a pit and the Wumpus, or a pit and the gold, or the Wumpus and the gold. The Wumpus does not move and the agent does not have ‘arrows’ so that she can kill the Wumpus.

<sup>12</sup> “Epistemic decision theory [EUT] usually presupposes that the credence functions it compares are defined over the same algebra of propositions [agenda]. Once we abandon this presupposition, new difficulties arise” (Carr 2015, 223). I exploit some of these difficulties in the rest of this paragraph.



the agent. The scoring rules in Table 1 encode different attitudes towards epistemic risk (see Babic 2019). The Brier score awards risk-averse agents who hold middling credences because  $f_\varphi = (\varepsilon_\varphi)^2 \leq \varepsilon_\varphi$ , where this difference is larger when  $\varepsilon_\varphi$  is around 0.5. The log score invites risk-seeking agents because  $t_\varphi = -\ln(\varepsilon_\varphi) \geq (1 - \varepsilon_\varphi)$  and  $f_\varphi = -\ln(1 - \varepsilon_\varphi) \geq \varepsilon_\varphi$ , where these differences approach  $+\infty$  when  $\varepsilon_\varphi$  approaches 0 and 1 (respectively, but see fn. 12). The linear score is neutral in this regard because  $t_\varphi = (1 - \varepsilon_\varphi)$  and  $f_\varphi = \varepsilon_\varphi$ . The Brier score tends to favor the Bayesian agent, who can hold credences. The log score may favor the nonmonotonic agent, who can hold all-or-nothing true beliefs and withhold those that are possibly false, or the Bayesian, who can avoid the risk by holding non-extreme credences. I will use the three scores to keep the results of the simulations independent of specific measures.<sup>13</sup>

The alethic function  $a(t, f)$  must strictly increase with respect to (wrt)  $t$  (if  $t' > t$ , then  $a(t', f) > a(t, f)$ ) and strictly decrease wrt  $f$  (if  $f' > f$ , then  $a(t, f') < a(t, f)$ ). These ‘basic requirements’ are accepted by Douven (2013, 436): “The basic intuition underlying it is clear enough, to wit, that the higher one’s degree of belief in a true proposition is, the more accurate one is, *ceteris paribus*, and also the lower one’s degree of belief in a false proposition is, the more accurate one is, *ceteris paribus*”. The basic requirements are put to work in a computational investigation by Trpin and Pellert (2019), who use the function  $t - f$ . This is the ‘minimal’ function that fulfills the basic requirements, but it is unbounded from above and below. This feature makes the comparison of performances in caves of different sizes difficult because the final value may be dominated by the values of  $t$  or  $f$ . This problem may be avoided by adding a denominator to this function, which may be done conscientiously in at least two ways: (i)  $t + f$ , as the agent’s ‘amount of commitment’ (there is a vestigial problem with this suggestion, see below), or (ii)  $n$ , as the maximum size for the agent’s agenda (the number of propositions at issue).<sup>14</sup>

---

<sup>13</sup> The Brier score is the most popular scoring rule in EUT, but Lewis and Fallis (2019) argue for the log score. The linear score is often dismissed for not being a proper scoring rule (see Lewis and Fallis 2019, sec. 4), but it returns the same values as the Brier score for full beliefs ( $0^2 = 0$  and  $1^2 = 1$ ). The fact that  $t_\varphi$  and  $f_\varphi$  collapse to  $+\infty$  when  $\varepsilon_\varphi$  is 0 and 1 causes the global measure of  $t$  and  $f$  also to collapse in those cases. To avoid this, I will compute  $\ln(c)$  instead of  $\ln(0)$  for a small constant  $c > 0$  (I will use  $c = 0.01$ , but there is nothing special with this value).

<sup>14</sup> The number of propositions in the real world is infinite, but it may be finite in toy worlds, such as EWW. It is often assumed that agendas are maximal (i.e., that agents are interested in the truth-value of ‘every proposition’). This is a reasonable idealization for our purposes, but real agents may withdraw propositions from their agendas (see the ‘anti-interrogative attitude’ in sec. 3).

The first possibility results in the function  $(t - f)/(t + f)$ , which does not fulfill the basic requirements and does not measure epistemic value correctly. For example, two agents with only true beliefs about one and ten propositions (respectively) would receive the same evaluation, but the second is more well-informed than the first and should be awarded for this. The addition of a small constant  $c > 0$  to the denominator avoids this problem,<sup>15</sup> leading to the second agent receiving more (marginal) epistemic value than the first. The resulting function is  $\alpha(t, f) = (t - f)/(t + f + c)$ , which is discussed in Dantas (2021). The second possibility results in the function  $\beta(t, f) = (t - f)/n$ , where  $n$  is the number of propositions at issue (in EWW,  $n = 2 \times s^2$ ). This function is used in the literature about truthlikeness (e.g., Cevolani and Festa 2021, 11472). The function  $\alpha$  awards lower amounts of commitment, which both agents can do: the Bayesian agent can hold middling credences (which lowers  $t + f$  given some scores) and the nonmonotonic agent can commit herself to fewer propositions (which always lowers  $t + f$ ). The denominator of the function  $\beta$  denominator is a fixed upper bound for the agents' amount of commitment and it does not award lower amounts of it. I will use both functions to keep the results of the simulations independent of specific measures. For readability, I will measure the epistemic value as  $1000 \times a(t, f)$ , where  $a = \alpha$  ( $\alpha$ -value) or  $a = \beta$  ( $\beta$ -value). The practical value (p-value) will be measured as +1000 for escaping with the gold and -1000 for dying.

### 1.1 The Bayesian model

The Bayesian model of a rational agent has the following features:

- b1. The agent's belief-values can have continuously many values between 0 and 1 (credences);
- b2. The agent's credence function is consistent with the axioms of probability;
- b3. The agent updates her credences using conditionalization.

The Bayesian agent holds very fine-grained beliefs (b1) and fulfills both probabilism (b2) and conditionalization (b3). It follows from b2 and b3 that the Bayesian agent is opinionated.<sup>16</sup> This model does not determine the initial

---

<sup>15</sup> I intend to set  $c = 0.01$  in both uses of constants (see fn. 12), but nothing in this paper depends on this specific value. The constant  $c$  may be seen as setting the 'sensitivity' of the function  $\alpha$ : the smaller the  $c$ , the greater the benefit for believing more truths and the penalty for believing more falsehoods.

<sup>16</sup> A (probabilistic) credence *function* (not a partial function) returns a value between 0 and 1 for every proposition in its domain (the agent's agenda) and there is no update from the absence of values about a proposition to some belief-value using conditionalization. As a result, the Bayesian agent cannot gain or lose beliefs.

credences of the Bayesian agent in the EWW (apart from their being probabilistic), but I will freely use principles such as the principal principle and the principle of indifference to determine those priors.

The following sentence schemas describe the agent's initial unconditional credences about pits and the Wumpus:

b4.  $cr(p_{x,y}) = pr(p)$  if  $(x, y) \neq (0, 0)$ ;

b5.  $cr(w_{x,y}) = 1/(s^2 - 1)$  if  $(x, y) \neq (0, 0)$ ;

b6.  $cr(p_{0,0}) = cr(w_{0,0}) = 0$ .

Clauses b4 state that the agent has an initial credence of  $pr(p)$  that  $(x, y)$  contains a pit for every room  $(x, y) \neq (0, 0)$ . Clauses b5 state that, for every room  $(x, y) \neq (0, 0)$ , the agent has initial credence of  $1/(s^2 - 1)$  that  $(x, y)$  contains the Wumpus. Clauses b6 state that the agent has an initial credence of 0 that  $(0, 0)$  contains a pit or the Wumpus.

The agent's conditional credences about pits and the Wumpus are:

b7.  $cr(b_{x,y} | \bigvee p_{\pm x, \pm y}) = 1$  and  $cr(b_{x,y} | \bigwedge \neg p_{\pm x, \pm y}) = pr(rand)$ ;

b8.  $cr(s_{x,y} | \bigvee w_{\pm x, \pm y}) = 1$  and  $cr(s_{x,y} | \bigwedge \neg w_{\pm x, \pm y}) = pr(rand)$ ;

where  $\bigvee \varphi_{\pm x, \pm y}$  abbreviates  $\varphi_{x+1,y} \vee \varphi_{x-1,y} \vee \varphi_{x,y+1} \vee \varphi_{x,y-1}$  and  $\bigwedge \neg \varphi_{\pm x, \pm y}$  abbreviates  $\neg \varphi_{x+1,y} \wedge \neg \varphi_{x-1,y} \wedge \neg \varphi_{x,y+1} \wedge \neg \varphi_{x,y-1}$ .<sup>17</sup> Clauses b7 and b8 are a consequence of the fact that if a room is in the neighborhood of a room containing a pit or the Wumpus, then it contains a breeze or a stench with probability 1; else it contains a random breeze or a random stench with probability  $pr(rand)$ . The Bayesian agent updates her beliefs using conditionalization.

## 1.2 The nonmonotonic model

The nonmonotonic model of a rational agent has the following features:

- d1. The agent's beliefs have only two potential values: 0, 1 (full beliefs);
- d2. The agent can withhold beliefs (*null* values);
- d3. The agent adopts and withdraws beliefs given adequate reasons and defeaters.

The nonmonotonic agent has doxastic states much less fine-grained than the Bayesian. Instead of continuously many values between 0 and 1, the model only allows for two belief-values (d1): 0 when the agent disbelieves that  $\varphi$  (i.e., she

---

<sup>17</sup> The Bayesian agent may not hold credences about complex propositions in EWW. In this case,  $\bigwedge$  and  $\bigvee$  should be seen as meta-linguistic connectives, where  $\bigwedge \neg \varphi_{\pm x, \pm y}$  abbreviates the list where  $\varphi_{x+1,y}$ ,  $\varphi_{x-1,y}$ ,  $\varphi_{x,y+1}$ , and  $\varphi_{x,y-1}$  appear all negated and  $\bigvee \varphi_{\pm x, \pm y}$  abbreviates the lists of these same atoms where at least one of them appears in the affirmative form (in the same order).

believes that  $\neg\varphi$  and 1 when she believes that  $\varphi$ . Another difference is that, while the Bayesian is opinionated, the nonmonotonic agent can withhold beliefs (d2). The withholding of belief is modeled by using a third value (*null*), which marks the absence of belief-values for some proposition  $\varphi$  (i.e.,  $\varphi$  is not in the agent's belief-set). Nonmonotonic frameworks are used to model defeasible reasoning (d3), where agents draw and retract conclusions given new information.

The nonmonotonic agent fulfills neither probabilism nor conditionalization as she initially holds *null* values about many propositions and gains new beliefs from investigating the cave (see fn. 15). LP is a nonmonotonic framework with computational applications (Doets 1994). A 'logic program' in LP is a set of conditionals of the form  $\varphi \wedge \neg ab \rightarrow \psi$  ('if  $\varphi$  and nothing is abnormal, then  $\psi$ '), where  $\varphi$  is a conjunction of literals,  $ab$  is a disjunction of literals ('abnormalities') indexed to a particular conditional, and  $\psi$  is a single literal. These conditionals may be seen as licenses for performing inferences (e.g., modus ponens) under certain conditions (e.g.,  $ab$  is false). If something is abnormal (i.e.,  $ab$  is true), then  $\varphi \wedge \neg ab$  is false and the modus ponens is blocked. If  $\varphi$  is true but  $\psi$  is false, then something is abnormal ( $ab$  is true). This framework exhibits a form of negation as failure (the closed world assumption): in some cases, the absence of reasons for the truth of some proposition (e.g.,  $ab$ ) is a reason for its negation. LP may be used to model defeasible reasoning<sup>18</sup>.

The nonmonotonic agent is modeled using LP. Her initial beliefs about (0, 0) are  $\neg p_{0,0}$  and  $\neg w_{0,0}$ . For all  $(x, y) \neq (0, 0)$ ,  $p_{x,y}$  and  $w_{x,y}$  are initially assigned the *null* value. The agent uses the following conditionals about pits:

$$d4. b_{x,y} \wedge \neg ab \rightarrow p_{\pm x, \pm y};$$

$$d5. \neg b_{x,y} \rightarrow \neg p_{\pm x, \pm y},$$

where  $\varphi_{\pm x, \pm y}$  is either  $\varphi_{x+1,y}$ ,  $\varphi_{x-1,y}$ ,  $\varphi_{x,y+1}$ , or  $\varphi_{x,y-1}$ . Clauses d4 state that perceiving a breeze in a room is a defeasible reason for believing that there are pits in the adjacent rooms, whereas  $ab$  expresses the possibility that  $b_{x,y}$  is a random breeze. Clauses d5 state that perceiving the absence of breezes in a room is a conclusive reason for believing that there aren't pits in the adjacent rooms. Clauses d5 express rebutting defeaters for the belief that  $p_{x,y}$  based on d4.

---

<sup>18</sup> If  $\varphi \wedge \neg ab \rightarrow \psi$  is a clause in the agent's logic program, then believing that  $\varphi$  is a reason for believing that  $\psi$ . If  $ab$  has zero disjuncts, then this is a conclusive reason. Otherwise, it is a defeasible reason. We could say that if  $\varphi'$  is a reason for  $\neg\psi$ , then the belief that  $\varphi'$  rebuts the belief that  $\psi$  on the previous basis; and that if  $\varphi'$  is a reason for  $ab$ , then the belief that  $\varphi'$  undercuts the belief that  $\varphi$  on that same basis. LP often blurs the distinction between rebutting and undercutting defeaters because a reason for  $\neg\psi$  is often a reason for  $ab$  (by modus tollens) and a reason for  $ab$  may be a reason for  $\neg\psi$  (closed world).

The agent uses the following conditionals about the Wumpus:

- d6.  $s_{x,y} \wedge \neg ab \rightarrow w_{\pm x, \pm y}$ ;
- d7.  $s_{x,y} \wedge \neg ab \rightarrow \neg w_{z,w}$  for  $(z, w) \neq (\pm x, \pm y)$ ;
- d8.  $w_{x,y} \rightarrow \neg w_{z,w}$  for  $(z, w) \neq (x, y)$ ;
- d9.  $\neg s_{x,y} \rightarrow \neg w_{\pm x, \pm y}$ ,

where  $(z, w) \neq (\pm x, \pm y)$  abbreviates  $(z, w) \neq (x + 1, y) \wedge (z, w) \neq (x - 1, y) \wedge (z, w) \neq (x, y + 1) \wedge (z, w) \neq (x, y - 1)$ . Clauses d6 state that perceiving a stench in a room is a defeasible reason for believing that there are Wumpuses in each of the adjacent rooms, whereas  $ab$  expresses the possibility of  $s_{x,y}$  being a random stench. Clauses d7 state that perceiving a stench in a room is a defeasible reason for believing that the Wumpus is nowhere else but in the adjacent rooms, where  $ab$  has the same interpretation. Clauses d6 and d7 may rebut each other. Clauses d8 are a consequence of the fact that there is only one Wumpus: believing that the Wumpus is in a specific room is a conclusive reason for believing that it is not elsewhere. Clauses d9 state that perceiving the absence of stench in a room is a conclusive reason for believing that the Wumpus is not in the adjacent rooms.

Integrating these conditionals in a sound pattern of inference is not trivial. Some researchers have proposed procedural semantics for LP, which are based on the fixpoints of an update operator.<sup>19</sup> For example, given a belief-set  $M$  and a logic program  $P$ , Stenning and van Lambalgen (2008, 194) define an application of the update operator  $T$  to  $M$  (i.e.,  $T(M)$ ) as:

- (a)  $T(M)(\psi) = 1$  iff there is a clause  $\varphi \rightarrow \psi$  in  $P$  such that  $M \models \varphi$ ;
- (b)  $T(M)(\psi) = 0$  iff there is a clause  $\varphi \rightarrow \psi$  in  $P$  and for all such clauses,  $M \models \neg \varphi$ ;
- (c)  $T(M)(\psi) = u$  otherwise (Stenning and van Lambalgen 2008, 194),

where their third value,  $u$  (currently indeterminate), is related to my use of the *null* value. The updated belief-set  $T(M)$  must be among the fixpoints of  $T$  (i.e.,  $T(T(M)) = T(M)$ ).<sup>20</sup> I will focus on the ‘most well-supported’ fixpoint, where a belief-set is more well-supported than another iff it is constructed by triggering more conditionals d6.<sup>21</sup>

<sup>19</sup> A fixpoint  $c$  of a function  $f(x)$  is such that  $c$  belongs to the domain and the codomain of  $f(x)$ , and  $f(c) = c$ .

<sup>20</sup> There may be different applications of  $T$  to a model  $M$  that are fixpoints of  $T$ . Suppose that  $M \models s_{0,1}, s_{2,1}$ . There are at least three applications  $T(M)$  that are fixpoints of  $T$ : (1)  $T(M)(w_{0,2}) = u$ ,  $T(M)(w_{2,2}) = u$ ; (2)  $T(M)(w_{0,2}) = 1$ ,  $T(M)(w_{2,2}) = 0$ ; and (3)  $T(M)(w_{0,2}) = 0$ ,  $T(M)(w_{2,2}) = 1$ .

<sup>21</sup> Suppose that  $M \models s_{0,1}, s_{1,2}, s_{3,1}, \neg s_{0,0}, \neg s_{1,0}, \neg s_{1,1}, \neg s_{2,0}, \neg s_{2,1}, \neg s_{2,2}, \neg s_{3,1}$ . In this case, the most well-supported model is one such that  $T(M)(w_{0,2}) = 1$  and  $T(M)(w_{3,2}) = 0$  because it triggers two conditionals d6 ( $s_{0,1} \wedge \neg ab \rightarrow w_{0,2}$  and  $s_{1,2} \wedge \neg ab \rightarrow w_{0,2}$ ) while the other models trigger at most one.

## 2. The simulation

I have implemented AI agents based on the Bayesian and nonmonotonic models in a computer simulation of EWW. The agents have four modules: perception, memory, practical and epistemic cognition. They share perception, memory, and part of practical cognition, but they have different epistemic cognitions (based on the Bayesian and nonmonotonic models respectively). The modules of perception and memory are straightforward. Perception simply receives percepts from the environment (e.g., the presence of a breeze), encodes them in the format of beliefs, and sends these beliefs to memory. Beliefs are formatted as one matrix of numbers for each of breeze, stench, pit, Wumpus, and gold, where a position in a matrix represents the corresponding room in the cave.<sup>22</sup> Memory simply stores these matrices and makes them available for practical and epistemic cognition.

The role of practical cognition is to construct and execute plans (sequences of actions). Practical cognition checks whether there is a plan being executed. If there is, it executes the next action in the plan. Otherwise, it constructs a plan and executes its first action. In constructing a plan, practical cognition chooses the sub-goal that she attributes higher value among ‘grab the gold,’ ‘get out of the cave,’ and ‘move to a fringe room  $(x, y)$ ,’ where fringe rooms are non-visited rooms in the neighborhood of a visited room. The value of moving to a room  $(x, y)$  is calculated differently by each agent.<sup>23</sup> The value of grabbing the gold (if the room contains it) is 1000 and the value of getting out of the cave is 0. The plan for moving to a room is constructed using Dijkstra’s pathfinding algorithm, with cost computed as the number of actions (Russell and Norvig 2020, 84); the plan for grabbing the gold is composed solely of the action ‘grab the gold;’ the plan for getting out of the cave is a plan for moving to  $(0, 0)$  concatenated with the action ‘get out of the cave.’

Epistemic cognition updates the agent’s beliefs, given the new information. The agents differ in how they encode and update beliefs about pits and Wumpus.<sup>24</sup> The Bayesian agent encodes those beliefs using floating-point numbers between 0 and 1 and updates them using causal Bayesian networks and joint probability

---

<sup>22</sup> For example, the position  $(x, y)$  in the pit matrix represents  $p_{x,y}$ , where the value in that position represents the agent’s belief-value for  $p_{x,y}$ .

<sup>23</sup> The Bayesian agent calculates the expected utility of moving to a room  $(x, y)$  as  $-1000 \times (cr(p_{x,y}) + cr(w_{x,y})) + 1000 \times cr(g_{x,y})$ . The nonmonotonic agent assigns  $-1000$  if  $b(p_{x,y}) = 1$  or  $b(w_{x,y}) = 1$ ; she assigns  $1000 \times b(g_{x,y})$  if  $b(g_{x,y}) \neq null$ ; and an intermediate value (e.g., 500) otherwise. This is an application of the minimax principle.

<sup>24</sup> For simplicity, both agents encode and update beliefs about breezes, stenches, and gold in the same all-or-nothing way: if the agent perceives a breeze [stench, gold] in a room, then perception adds a 1 to the corresponding position in the breeze [stench and gold] matrix; otherwise, it adds a 0.

tables, in a standard algorithm (e.g., Russell and Norvig 2020, 134). The nonmonotonic agent encodes beliefs about pits and Wumpus using Boolean values. The algorithm for pits is very simple. If you perceive a breeze in a position  $(x, y)$ , then write a 1 in all positions  $(\pm x, \pm y)$  of the pit matrix that do not already contain a 0 (d4). If you perceive the absence of breezes in a position  $(x, y)$ , then write a 0 in all positions  $(\pm x, \pm y)$  of the pit matrix (d5). The algorithm for Wumpus keeps track of its viable locations and the number of cues about the Wumpus. Do the following when you perceive a new stench. If the stench matrix contains a 1 in a position  $(x, y)$ , then add +1 to the value in the positions  $(\pm x, \pm y)$  of the support matrix (d6) and add -1 from the value in the other (d7 and d8). If the stench matrix contains a 0 in a position  $(x, y)$ , then write a large negative number in the positions  $(\pm x, \pm y)$  of the support matrix (d9). If there is a position in the support matrix with a value that is strictly higher than the other, assign 1 to that position in the Wumpus matrix and 0 to the other. Otherwise, assign *null* to the positions with the highest value and 0 to the other.

The simulation of EWW generates a random cave for each trial and works as a loop. In each iteration, the cave outputs precepts for the agent, who updates her beliefs and returns an action to the cave, which updates its state given that action. The loop stops when the agent dies or gets out of the cave. I have simulated three conditions (c1-c3). In c1, the probability of pits is fixed at 0.1 ( $pr(p) = 0.1$ ) and the probability of random breezes [stenches] is fixed at 0.01 ( $pr(rand) = 0.01$ ). I have run the simulation using caves with dimensions  $s \times s$  (size  $s$ ), from  $s = 2$  to  $s = 10$  (incrementing by 1). In c2, I have run simulations with  $pr(rand)$  varying from 0 to 0.5 (incrementing by 0.05), where  $pr(p)$  is fixed at 0.1. In c3, I have run simulations with  $pr(p)$  varying from 0 to 1 (incrementing by 0.1), where  $pr(rand)$  is fixed at 0.01. In c2 and c3, I have used ‘medium’ caves ( $s = 6$ ). I have run up to 50,000 trials for each configuration of the cave in each condition and averaged the results. The code was written in Java and the graphs were plotted in Grace.

## 2.1 Results

The results for the Bayesian and nonmonotonic agents are depicted in Figures 1 and 2 (respectively). The linear and the Brier scores return the same  $\beta$ -values for both agents because  $t - f$  returns the same values for these scores and  $n$  is constant within each trial.<sup>25</sup> The linear and the Brier scores do not return the same  $\alpha$ -values for the Bayesian agent because  $t + f$  does not return the same values for these

---

<sup>25</sup> Let  $y = \varepsilon_\varphi$  and  $x = 1 - \varepsilon_\varphi$ . The function  $t - f$  returns the same values for both agents in terms of linear and Brier scores because  $(x + y) \times (x - y) = x^2 - y^2$  and  $x + y = 1$ , which entails that  $x - y = x^2 - y^2$ .

scores, but this is the case for the nonmonotonic because  $0 = 0^2$  and  $1 = 1^2$ . The log values were normalized by  $-\ln(c)$ , as to depict all scores in the same scale.<sup>26</sup> In this case, the three scores return the same  $\alpha$ - and  $\beta$ -values for the nonmonotonic agent because  $1 = 1^2 = (-\ln(c))/(-\ln(c))$  and  $0 = 0^2 = (-\ln(1))/(-\ln(c))$ . I will discuss  $f$  and the time requirements in Section 3.

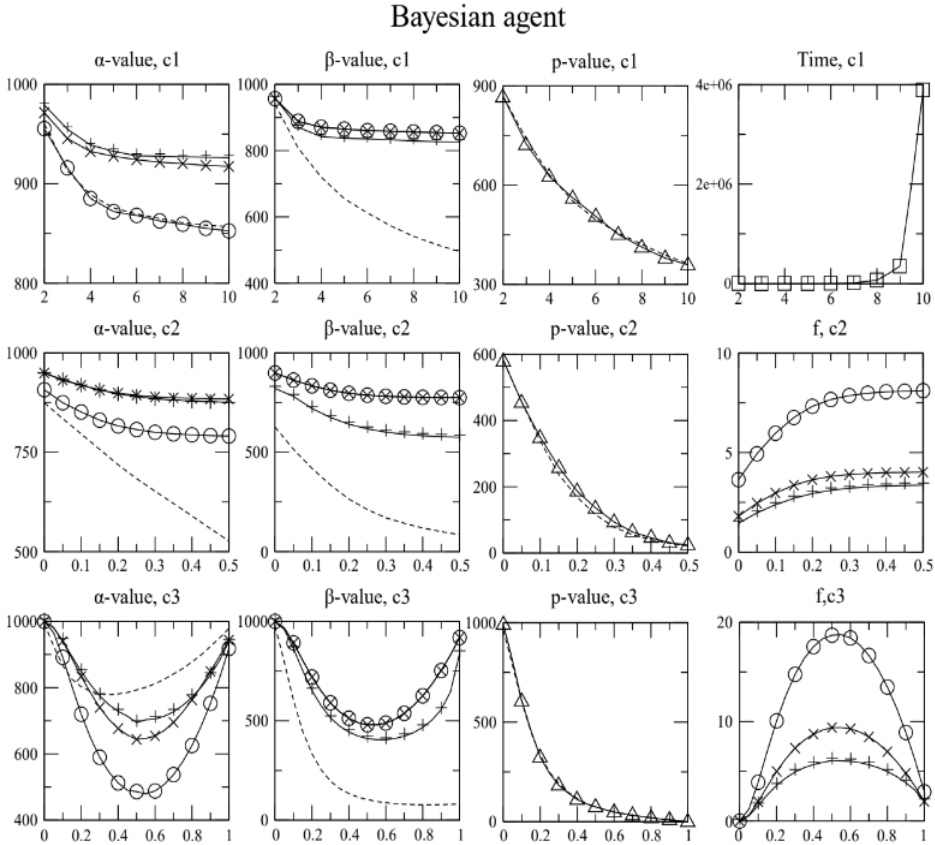


Fig. 1: Top to bottom: results of the Bayesian agent for c1, c2, and c3. Left to right: the  $\alpha$ - and  $\beta$ -values, using linear ( $\circ$ ), Brier ( $\times$ ), and log scores ( $+$ ); and the p-value ( $\Delta$ ). The results for the nonmonotonic agent are depicted in dashed lines. The 4th column depicts the time requirements (number of variable-changes) in c1 ( $\square$ ) and  $f$  in c2 and c3 using the same scoring rules.

Condition c1 represents the ‘normal’ condition, in the sense of the evidence being relatively informative and relatively accessible for the agents. In c1, as the cave

<sup>26</sup> This was unnecessary for the  $\alpha$ -values because the divisor  $t + f + c$  already works as a normalizer.



grows from  $s = 2$  to  $s = 10$ , the  $\alpha$ -values of the Bayesian agent start stabilizing around 853 (linear), 917 (Brier), or 926 (log). The  $\alpha$ -values of the nonmonotonic agent start stabilizing around 857. The Bayesian's  $\alpha$ -values are higher than the nonmonotonic's when the measure uses a Brier or log score, but they are not when it uses a linear score (they are about the same, with similar curves). The  $\beta$ -values of the Bayesian agent start stabilizing around 852 (linear and Brier), or 827 (normalized log). The  $\beta$ -values of the nonmonotonic agent go down to 495 and start to stabilize at a much lower level (250).<sup>27</sup> The Bayesian's  $\beta$ -values are higher than the nonmonotonic's independently of the scores. The results of  $\alpha$ - and  $\beta$ -values diverge, as the nonmonotonic agent is much closer to the Bayesian in terms of  $\alpha$ -values (especially, in linear score) than she is in terms of  $\beta$ -values. However, the disagreement is not very strong as all graphs have the same general shape: they fall fast until stabilizing at different levels. Furthermore, the (higher) epistemic value of the Bayesian agent does not reflect in a higher practical value: both agents have p-values that reach around 360 when the cave reaches  $s = 10$  (with very similar curves).

In c2, I have tested how the agents react when  $pr(rand)$  varies from 0 to 0.5 (the tendencies do not change from 0.5 to 1). The higher the  $pr(rand)$ , the less information observed breezes [stenches] carry about the position of pits [Wumpus].<sup>28</sup> The  $\alpha$ -values of the Bayesian agent stabilize around 775 (linear), 874 (Brier), or 860 (log) when  $pr(rand)$ . The nonmonotonic agent performs much worse: her  $\alpha$ -values reach 527 when  $pr(rand) = 0.5$ . The same goes for  $\beta$ -values. The Bayesian's  $\beta$ -values stabilize around 775 (linear and Brier), or 575 (normalized log), while the nonmonotonic's fall to 86 when  $pr(rand) = 0.5$ . These results highlight the strength of the Bayesian model: the ability to deal with uncertain evidence. The Bayesian agent considers the value of  $pr(rand)$  in updating beliefs, whereas the nonmonotonic can only conclude blindly that there are pits (not so much for Wumpus) in the surroundings from the presence of breezes (random or not). The disagreement between the  $\alpha$ - and  $\beta$ -values is also mild in c2, where the nonmonotonic's  $\beta$ -values fall faster than her  $\alpha$ -values. The higher epistemic value of the Bayesian agent affects her practical value: although both p-values converge to around 22 when  $pr(rand) =$

---

<sup>27</sup> The  $\beta$ -values of the nonmonotonic agent are around 250 from  $s = 50$  to  $s = 150$  (her  $\alpha$ -values are around 845 in those cases). These results are less reliable because they were averaged over fewer trials.

<sup>28</sup> For example, suppose that the agent has just arrived in the cave. If  $pr(p) = 0.1$  and  $pr(rand) = 0.01$ , then conclusive evidence about  $b_{0,1}$  transmits 0.27 bits of information about  $p_{0,2}$ . If  $pr(p) = 0.1$  and  $pr(rand) = 0.05$ , then conclusive evidence about  $b_{0,1}$  transmits only 0.24 bits of information about  $p_{0,2}$ .

0.5, the Bayesian's p-value (94) is 20% higher than the nonmonotonic's (78) when  $pr(rand) = 0.3$ .

Nonmonotonic agent

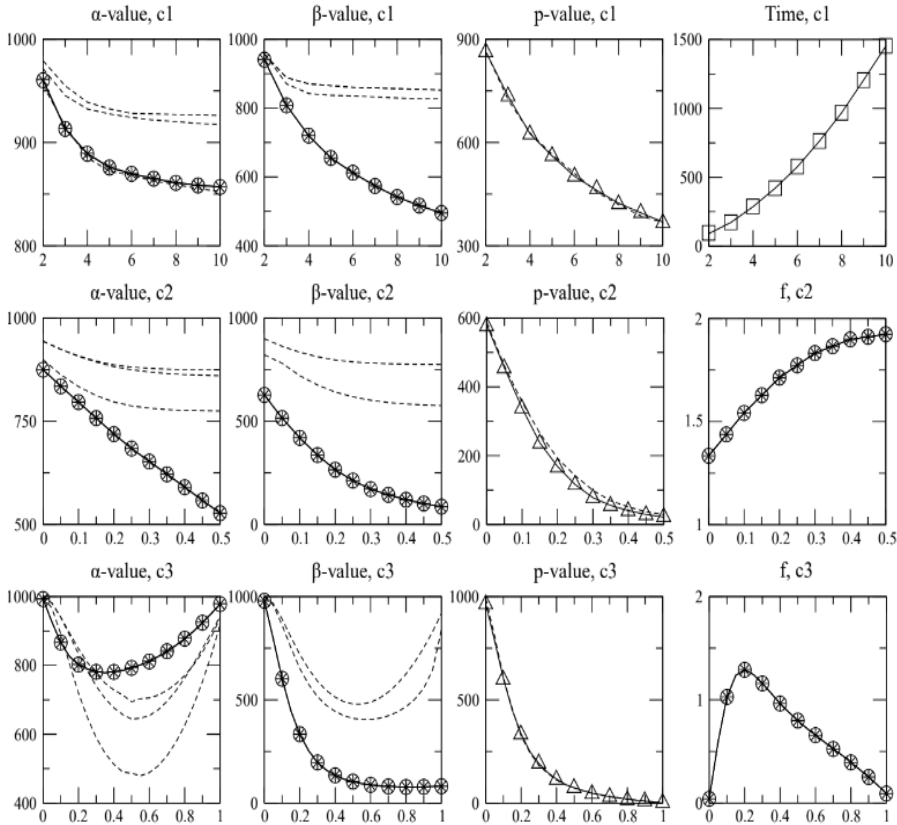


Fig. 2: Top to bottom: results of the nonmonotonic agent for c1, c2, and c3. Left to right: the  $\alpha$ - and  $\beta$ -values, using linear ( $\circ$ ), Brier ( $\times$ ), and log scores ( $+$ ); and the p-value ( $\Delta$ ). The results for the Bayesian agent are depicted in dashed lines. The 4th column depicts the time requirements (number of variable-changes) in c1 ( $\square$ ) and  $f$  in c2 and c3 using the same scoring rules.

In c3, I have tested how the agents react when  $pr(p)$  varies from 0 to 1. The increase in  $pr(p)$  increases the difficulty of gathering the evidence because the higher the number of pits, the riskier it is to explore the cave. The  $\alpha$ -values of the Bayesian reach around 918 (linear), 944 (Brier), or 939 (log) when  $pr(p) = 1$ , in a u-shape with the lowest values around 479 (linear), 644 (Brier), or 695 (log) when

$pr(p) = 0.5$ . The nonmonotonic's  $\alpha$ -values reach around 978 when  $pr(p) = 1$ , in a u-shape with the lowest values around  $pr(p) = 0.35$  (778). The nonmonotonic agent performs much better than the Bayesian in terms of  $\alpha$ -values when  $pr(p) > 0.35$ , which is surprising because the Bayesian (but not the nonmonotonic) agent fulfills probabilism and conditionalization. The explanation for this result is that it is too risky to gather information when  $pr(p) > 0.35$  when the nonmonotonic's *null* values about the unobserved rooms are worth more  $\alpha$ -value than the Bayesian's indifferent beliefs.<sup>29</sup> The situation is not the same regarding  $\beta$ -values. The  $\beta$ -values of the Bayesian reach around 918 (linear and Brier) or 839 (normalized log) when  $pr(p) = 1$ , in a u-shape with the lowest values around 479 (linear and Brier) and 404 (normalized log) when  $pr(p) = 0.5$ . The nonmonotonic's  $\beta$ -values do not have a u-shape, as they fall to 80 when  $pr(p) = 1$ . Neither the nonmonotonic's higher  $\alpha$ -values nor the Bayesian's higher  $\beta$ -values affect their p-values (both fall to 0 with similar curves).

In c1 and c2, the functions  $\alpha$  and  $\beta$  only disagree on the speed at which the curves decrease and at which level they converge. In c3, they disagree on the very shape of the graphs: the  $\alpha$ - but not the  $\beta$ -values of the nonmonotonic agent have a u-shape. They also disagree on which agent achieves higher epistemic value: the  $\alpha$ -values of the nonmonotonic agent are higher than the Bayesian's when  $pr(p) > 0.35$ , but the  $\beta$ -values of the Bayesian are always higher than the nonmonotonic's. These results are related to the relative accessibility of evidence in c1-c3. The functions  $\alpha$  and  $\beta$  generally agree in c1 and c2, where the evidence is accessible, independently of whether it is informative (c1) or misleading (c2). They disagree in c3, where the evidence is increasingly inaccessible when  $pr(p) > 0.35$ . Consequently, the functions  $\alpha$  and  $\beta$  generally agree on their evaluations of how agents reason from the available evidence (c1 and c2), but they disagree about what they should do in the absence of evidence (c3). In the absence of evidence, the agents tend to maintain their initial belief-values. The initial belief-values of the Bayesian agent (e.g., about the Wumpus) are often applications of the principle of indifference; the nonmonotonic's are mostly *null*. The use of *null* values may be interpreted as the withholding of beliefs. Then the disagreement between the

---

<sup>29</sup> Suppose that the Wumpus may be in three different rooms, but that the evidence about its position is inaccessible. The Bayesian agent would hold indifferent credences of 0.33 about these positions, while nonmonotonic would withhold beliefs (*null* values). The nonmonotonic agent would have a higher  $\alpha$ -value regarding those positions because  $\alpha(t, f) > \alpha(t + 1.67, f + 1.33)$  when  $t > f$ , which is usually the case in EWW. The Bayesian would have a higher  $\beta$ -value because  $\beta(t, f) < \beta(t + 1.67, f + 1.33)$ . These inequalities use linear scores, but the same holds for the Brier and log scores.

functions  $\alpha$  and  $\beta$  in c3 may be interpreted as they prescribing different attitudes in the absence of evidence: the function  $\alpha$  prescribes the withholding of beliefs; the  $\beta$  prescribes indifference (see fn. 28).

### 3. Discussion

The results from EWW may affect the justificatory force of the maximization arguments because the nonmonotonic agent (who fulfills neither probabilism nor conditionalization) achieves higher  $\alpha$ -values than the Bayesian in c3 (when  $pr(p) > 0.35$ ) and similar p-values in the three conditions. Whether this is the case depends on the adoption of the function  $\alpha$  as the measure of epistemic value and of a set of assumptions different from a1-a3. I will defend these choices, starting by proposing three arguments for the adoption of the function  $\alpha$ . The first argument is pragmatic. In c3, the  $\alpha$ -values of the nonmonotonic agent are higher than the Bayesian's when  $pr(p) > 0.35$ , but the  $\beta$ -values of the Bayesian are always higher than the nonmonotonic's. When  $pr(p) > 0.35$ , the Bayesian agent holds many more beliefs than the nonmonotonic (e.g., about the Wumpus). These 'extra beliefs' do not have practical value because the p-values of the Bayesian and nonmonotonic agents are similar in this condition. While the function  $\beta$  awards the Bayesian agent for holding beliefs without practical value, the function  $\alpha$  awards the nonmonotonic agent for not holding those beliefs. From a pragmatic point of view, the function  $\alpha$  is correct.<sup>30</sup> The pragmatic argument is controversial because it presupposes a form of pragmatic encroachment in the notion of epistemic rationality, whereas the epistemological orthodoxy most often separates epistemic rationality from practical rationality. Nevertheless, there are investigations that consider the possibility of pragmatic encroachment in the notion of epistemic rationality (e.g., Gao 2021).

The second argument is cognitive. A fundamental aspect of our cognitive situation is that human beings are in the finitary predicament of having fixed limits on their cognitive capacities and the time available to them (Cherniak 1986, 8). Epistemic rationality seems to require from 'finite reasoners' (those in the finitary predicament) a form of cognitive parsimony: to convert scarce cognitive resources (memory and time) into epistemic value efficiently. There are no interesting limits on the amount of information that we can hold in long-term memory (Dudai 1997), but the learning of new information can adversely affect our capacity to retrieve old information in a process of interference (Baddeley et al. 2020, 291).

---

<sup>30</sup> In c1, both agents are able to explore the cave and end up holding a similar number of beliefs. The Bayesian agents may hold extra beliefs in c2, but these beliefs do affect her p-value.

This cognitive limitation is modeled by adopting not an upper bound for the size of  $B$  but a diminishing reward for the amount of (truthful) commitment, which is done by the function  $\alpha$  because  $\alpha(t + 2x, f) - \alpha(t + x, f) < \alpha(t + x, f) - \alpha(t, f)$ , but not by the  $\beta$  because  $\beta(t + 2x, f) - \beta(t + x, f) = \beta(t + x, f) - \beta(t, f)$ , where  $x > 0$ . This difference between the functions  $\alpha$  and  $\beta$  explains their incompatible prescriptions in c3, where the nonmonotonic agent's withholding of beliefs in the absence of evidence is more cognitively parsimonious than the Bayesian's indifference (see Dantas 2022, for a discussion). If the notion of epistemic rationality should regard finite reasoners, then the function  $\alpha$  should be preferred over the function  $\beta$ . The Bayesian model is also more cognitively demanding in time (see g. 1 and 2, 1st line and 4th column).<sup>31</sup>

The third argument is epistemic, as it draws on veritistic notions. The function  $\alpha$  should be preferred over the function  $\beta$  because of how they relate to the most used measure of epistemic disvalue (inaccuracy). The goal of believing truths and avoiding falsehoods is two-fold, but I believe that its second part should take precedence over the first because of 'the problem of contradictory pairs.' The function  $\beta$  evaluates equally an agent who believes (as to the same degree) both propositions in a contradictory pair and one who believes neither because  $\beta(t, f) = \beta(t + x, f + x)$  when  $x > 0$ , but the second agent should be evaluated higher than the first regarding these propositions. This problem could be avoided by attributing weights  $R$  and  $W$  to  $t$  and  $f$  (respectively) where  $R < W$  (Fitelson and Easwaran 2015, 83), but this is to attribute priority to the goal of avoiding falsehoods. The function  $\alpha$  deals more naturally with this problem because  $\alpha(t, f) > \alpha(t + x, f + x)$  when  $x > 0$  and  $t > f$  but this is to attribute priority to the goal of avoiding falsehoods when  $t > f$ .<sup>32</sup> The nonmonotonic agent has a much lower inaccuracy than the Bayesian in c3, as depicted in figures 1 and 2, 2nd and 3rd lines, 4th

---

<sup>31</sup>Time requirements were measured as the number of changes in the value of variables used in update procedures. The time requirements of the nonmonotonic agent grow polynomially on the size of the cave; the Bayesian's grow exponentially (see fig. 1 and 2, 1st line and 4th column). There are algorithms for Bayesian inference that are polynomial in time (e.g., belief propagation, see Pearl 1986), but these only work for singly connected networks, whereas EWW requires multiple-connection. Inference from multiply connected Bayesian networks is NP-hard (Cooper 1990). Dantas (2017) argues that rationality demands finite reasoners to implement polynomial patterns of inference when they are available.

<sup>32</sup> The situation is the opposite when  $t < f$ , where the agent is an anti-expert about her agenda and believes a contradiction may serve as a flag for revising her beliefs (Dantas, 2022, discusses this point).

column.<sup>33</sup> If avoiding falsehoods takes precedence over believing truths, then the evaluation of the function  $\alpha$  in c3 should be preferred over that of the function  $\beta$ .

The function  $\beta$  awards investigation more straightforwardly than the function  $\alpha$  because the  $\beta$ -value, for example, of a new true (maximal) belief, is fixed ( $1/n$ ), but its  $\alpha$ -value varies inversely wrt the overall truthfulness of the agent's belief-set. If the agent holds mostly true beliefs (and many of them), the  $\alpha$ -value of investigating will be only marginal. This feature of the function  $\beta$  seems welcome, but I believe that the way that the function  $\alpha$  awards investigation is more appropriate for c3. Figures 1 and 2, 2nd and 3rd lines, 1st, 2nd, and 4th columns show that the graphs for  $\alpha$ -values have approximately the same shape as those of inaccuracy, although vertically reflected because  $\alpha$  is a value and inaccuracy a disvalue (the same holds in c1). This does not happen with  $\beta$ -values, especially in c3, where the graphs for  $\alpha$ -values and  $f$  have a u-shape for both the Bayesian and the nonmonotonic agents, but the graphs for  $\beta$ -values only have a u-shape for the Bayesian. This discrepancy is not welcome in c3, where the agents are not able to investigate, and avoiding error ( $f$ ) is even more important than seeking the truth ( $t$ ). The conservativeness of the function  $\alpha$  regarding the measure of inaccuracy seems to return the correct evaluation in c3. The functions  $\alpha$  and  $\beta$  agree in c1 and c2. If my arguments are good, the first should be preferred when they disagree (c3). The function  $\alpha$  should be preferred as a measure of epistemic value in general, especially regarding finite reasoners.

I have already commented on a1-a3 and EEW's alternative assumptions (a1' and a2'), but I will return to this point before addressing the DBAs and accuracy arguments. Assumption a1 is reasonable, but it is reasonable to risk value at the present if this may result in a gain of value in the future. This remark points to the reasonableness of a1', where the values are measured only at the end of each trial. 'Long run' evaluations of this sort may raise the concern that finite reasoners cannot always wait to fulfill their goals. They may also hide short run vices of the agents' reasoning processes. These concerns may be mitigated by paying attention to whether the agent's reasoning processes enable them to fulfill their goals. The practical performance of the Bayesian agent could be seen as the gold standard because she minimizes expected inaccuracy in every belief update (given a3) and

---

<sup>33</sup> This result does not contradict the conclusions of the accuracy arguments because the Bayesian and nonmonotonic agents are not opinionated over the same agenda. In c1 and c2, the nonmonotonic agent also has a lower inaccuracy, but both functions agree that the Bayesian achieves a higher epistemic value. In c1, as the cave grows from  $s = 2$  to  $s = 10$ , the nonmonotonic's inaccuracy grows from 0.1 to 3, whereas the Bayesian's grows from 0.2, 0.1, or 0.3 to 15, 7, or 27 (linear, Brier, and log scores respectively).

acts to maximize expected practical value (see fn. 22). The agents achieve similar p-values in the three conditions of EWW, which suggests that their reasoning processes are equally supporting the fulfillment of their goals (except in c2, where all measures show an unmistakable advantage for the Bayesian). These results suggest that both agents are able to fulfill their goals and the measuring of values only at the end of each trial in EWW is not hiding short run vices of the nonmonotonic agent. The related assumption that rationality requires agents not to make choices that lead to a sure loss of any practical value is also reasonable, but it depends on idealizing environments that do pose trade-offs between goals (I will return to this point in discussing the DBAs).

Assumption a2 is also reasonable because rationality is usually assumed to have a subjective character (e.g., Wedgwood 2015, 221), which is related to what the agent can know from the available evidence. Philosophers often distinguish between subjective and objective norms, where subjective but not objective norms are sensitive to which evidence is available to the agents. The veritistic norm is objective because the truth-values of their beliefs are often not transparent for the agents. This is why the Bayesian epistemologists supplement their measures of value with the notion of a sure loss or principles of decision theory, to consider rationality's subjective character. The resulting Bayesian norms are subjective. The expected values of EWW (the mean of actual values) are objective and I propose that their maximization is by itself relevant to the agent's rationality (this is the assumption a2'). This focus highlights interesting aspects of rationality. Together with a1', the objective expected values may be seen as evaluating the stable beliefs of agents, i.e., those that result from all the evidence available to her (including that available in the environment). The very notion of evidence being 'available' to a finite reasoner depends on the amount cognitive resources available to investigate. The evidence accessible through investigating the environment may also be 'available' in the relevant sense given a1' and a2'. This discussion suggests an ecological notion of rationality (Todd and Gigerenzer 2007), where the rationality of an agent depends on how she copes with her surroundings.

If a1 and a2 are reasonable assumptions, a3 is not. Opinionation is assumed as a simplifying idealization in EUT because of the limitations of the measure of inaccuracy (see sec. 1), but the results of EWW suggest that non-opinionated epistemic practices may be worth more epistemic value ( $\alpha$ -value) than opinionated ones (e.g., when evidence is not accessible). In this context, assuming a3 introduces a bias towards Bayesianism because it artificially eliminates these practices 'from the competition.' This assumption also defeats the purpose of veritism, of supporting norms of epistemic rationality from veritistic considerations. The

function  $\beta$  might be used for supporting opinionation from veritistic considerations. For example, indifference is worth more  $\beta$ -value than the withholding of belief when the agents lack evidence about large exhaustive sets of exclusive propositions (e.g., fn. 28).<sup>34</sup> Since the function  $\beta$  abstracts from the cognitive limitations of agents, its use for supporting opinionation would vindicate the Bayesian model as describing an ideal reasoner (i.e., a reasoner without cognitive limitations). This seems correct because the Bayesian ideal reasoner exhibits a form of logical omniscience<sup>35</sup> and cannot forget.<sup>36</sup>

Choices that lead to a sure loss should be avoided, but rational agents may be forced to make such choices when the environment poses trade-offs between goals. For example, Douven (2013) simulates a Bayesian and an explanationist agent, who updates her credences using inference to the best explanation and is vulnerable to DBs. These agents watch finite sequences of coin tosses and must estimate the coin's bias. Douven sets up a game, where the first correct estimation yields a point to the estimator, and an incorrect one yields a point to the opponent (the trade-off is between speed and accuracy). The Bayesian agent always loses because the explanationist converges to an answer faster. Part of the DBAs' appeal results from non-Bayesian agents being in the position to know that their choices lead to a sure loss in DB-environments. But the same is true about Douven's game because the Bayesian ideal reasoner is in the position to know anything that we can only learn from computer simulations (it is not surprising that slow convergence reasoning processes are prone to lose in environments that award speed of convergence). Then why would DB-environments be so central to rationality? I believe they are not, especially because Dutch bookies only occur as fictional characters in philosophers' tales (Douven 2013, 431). Although (ecological) rationality is relative

---

<sup>34</sup> This is not the case for exhaustive pairs of exclusive propositions (e.g., contradictory pairs) because  $\beta(t + x, f + x) = \beta(t, f)$ . How large these sets must be hangs on the relative sizes of the weights  $R < W$ , which could be adopted to deal with the problem of contradictory pairs (see the epistemic argument).

<sup>35</sup> The Bayesian ideal reasoner fulfills probabilism and conditionalization. The probability's axiom of normality entails that she must be certain of (i.e., hold maximum credence about) every logical truth. If she comes to learn some evidence with certainty, then the axioms of normality and finite additivity require her to be certain of every logical consequence of that evidence (see Garber 1984, 104).

<sup>36</sup> If the Bayesian ideal reasoner ever reaches certainty on a proposition, then the axioms of probability assure that this certainty will be maintained after any subsequent update by conditionalization. Consequently, the Bayesian ideal reasoner cannot be certain that she is currently having spaghetti for dinner, but forget this irrelevant fact a year later (i.e., loose certainty about it, see Talbott 1991, 139).



to sets of conditions (e.g., the Bayesian model copes better with misleading evidence), the relevance of these conditions depends on how likely they are to occur to reasoners like us (i.e., how ‘normal’ they are).

The results of EWW impact the accuracy arguments more than the DBAs because the nonmonotonic agent (who fulfills neither probabilism nor conditionalization) achieves higher  $\alpha$ -values than the Bayesian in c3 (when  $pr(p) > 0.35$ ). This result suggests that the Bayesian ideal reasoner is only guaranteed to maximize epistemic value in situations where a1-a3 hold. In those situations, she minimizes (expected) inaccuracy. But this result does not hold when assumption a3 is relaxed because the nonmonotonic agent achieves lower inaccuracy in the three tested conditions (this happens even in c2, in which the Bayesian model achieves higher  $\alpha$ - and  $\beta$ -values). The results of EWW impact more directly the argument for probabilism than that for conditionalization because, when  $pr(p)$  is high, the agents are not able to gather evidence and update beliefs (they tend to maintain their initial beliefs). However, the argument for conditionalization is also affected. The nonmonotonic’s higher  $\alpha$ -values were explained by her withholding of beliefs in the absence of evidence (see fn. 28) and there is no update from the absence of beliefs to a belief-value by conditionalization. An agent who maximizes  $\alpha$ -value in the absence of evidence cannot update beliefs by conditionalization. Conversely, the Bayesian ideal reasoner’s failure in maximizing  $\alpha$ -value was explained by her use of the principle of indifference, which is how she maintains opinionation in the absence of evidence. The inefficiency of the Bayesian model was located in its demand for opinionation (see fn. 15).

An epistemic interpretation of the nonmonotonic agent’s *null* values is that she is suspending judgment (i.e., adopting a neutral stance toward the truth-value of a proposition). A difficulty with this interpretation is that the mere lack of belief-values about a proposition is not sufficient for suspension.<sup>37</sup> There are different notions of suspension, two of which may be used to interpret the nonmonotonic’s use of *null* values: the interrogative view (Friedman 2015), in which suspending about a proposition involves actively inquiring about its truth, and the anti-interrogative view (Lord 2020), in which suspending about a proposition involves forgoing evidence about its truth. Intuitively, the nonmonotonic agent does the first when she attributes *null* values at the beginning of each task and when there is evidence of Wumpus in different rooms. She does the second at the end of each task when it is too risky to gather evidence about the

---

<sup>37</sup> For example, when an agent never considered a proposition, we say that she does not hold a doxastic attitude towards it and not that she holds an attitude of committed neutrality (suspension) towards it (Friedman 2013, 167).

remaining *null* propositions. There is a notion of suspension in which the Bayesian agent may be said to be suspending in c3: the credal view (Sturgeon 2008), where suspending is related to the holding of middling credences. The Bayesian agent holds middling credences about pits when  $pr(p) = 0.5$ , for example, because her initial (indifferent) credences are of 0.5 and it is difficult to gather evidence when  $pr(p) > 0.35$ . This form of suspension does not result in higher  $\alpha$ -values (but results in higher  $\beta$ -values). This is a matter for another paper.

#### 4. Conclusions

The conclusions of the maximization arguments follow from assumptions a1-a3, but the results of EWW show that these conclusions do not hold when an alternative set of reasonable assumptions is assumed. For example, the results from EWW show that there is a reasonable notion of maximization of epistemic value ( $\alpha$ -value, without the assumptions a1-a3) in which a nonmonotonic agent, who does not fulfill the Bayesian norms, maximizes epistemic value. The nonmonotonic agent also achieves the same amount of practical value in the ‘normal’ condition of c1, even though she is vulnerable to DBs. These results suggest that which assumptions and measures of epistemic value should be used in an investigation of rationality is open to discussion. I have argued that a1 and a2 are reasonable assumptions, but the alternative assumptions of EWW are also reasonable. The situation is different with a3 (opinionation), which, I have argued, is not a reasonable assumption. Researchers such as Gigerenzer and Gaissmaier (2011) have already investigated conditions under which Bayesian reasoning is epistemically outperformed by simple reasoning heuristics.<sup>38</sup> Their claims are corroborated by our results, especially because the nonmonotonic’s algorithm for Wumpus implements a tallying heuristic (Gigerenzer and Gaissmaier 2011, 469). The results of EWW are interesting because they locate the focus of epistemic inefficiency of the Bayesian model in its demand for opinionation. Forgoing this assumption is difficult for EUT because inaccuracy cannot be used to compare agents with different numbers of beliefs, which may be done with the functions  $\alpha$  and  $\beta$ .

The preceding discussion motivates the distinction between two different projects about rationality. The first is the Bayesian project, which relies on the assumptions a1-a3 and uses functions such as the function  $\beta$  for measuring

---

<sup>38</sup> “In a number of large worlds, simple heuristics were more accurate than standard statistical methods that have the same or more information. These results became known as ‘less-is-more effects’: there is an inverse-U-shaped relation between the level of accuracy and amount of information, computation, or time. There is a point where more is not better, but harmful” (Gigerenzer and Gaissmaier 2011, 453).

epistemic value. These assumptions enable the focus on analytical methods within this project (see below) because they introduce several simplifications. For example, the practical side of a1 requires idealizing that environments do not pose trade-offs between goals. These environments are ‘cost-free’ in the sense that no practical cost is associated with the gathering of evidence. This allows the Bayesian project to separate the epistemic from the practical. This project also abstracts from the cognitive limitations of finite reasoners because the function  $\beta$  does not consider these limitations (the value of new true beliefs does not depend on the agent’s belief-set). The assumption of opinionation consolidates this state of affairs because it forces the agents’ belief-sets to be fixed in size. The goal of the Bayesian project is to describe an ideal reasoner, but it might still concern finite reasoners (indirectly). The strategy would be to propose a model as an ideal reasoner whom we should strive to approximate (Leitgeb 2014, fn. 3). Approximating the Bayesian ideal reasoner would be beneficial for finite agents, even when they cannot be ‘fully’ rational. For example, De Bona and Stael (2017, 2018) show that, in doing so, finite reasoners are worth more epistemic value and become less vulnerable to DBs. The Bayesian project is an axiomatic approach to rationality, where optimization is related to approximating the model.

The second project is the one that I am proposing in this paper, which measures epistemic value using the function  $\alpha$  and drops the assumptions a1-a3. The goal of this project is to investigate the rationality of finite reasoners, where the function  $\alpha$  considers their cognitive limitations. The dropping of a1 and a2 has consequences for the methods of the investigation because these are the assumptions that enable the use of analytical methods in the Bayesian project. For example, a2 enables the consideration of all possible truth-values of the agent’s beliefs instead of their actual value. The study of actual values may hardly be carried out analytically. For example, the choice of a situation to be the actual (as to calculate actual values) would be unmotivated. This difficulty is avoided in computational epistemology, where the actual values are related to randomly generated environments. I believe that it might be a consequence of the no-free-lunch theorems (Wolpert and Macready 1997) that no general model of a rational agent will come out as maximizing practical or epistemic value in every environment. The rationality of a finite reasoner would depend not on how she approximates a model but on how she is able to exploit the features of her environment given her cognitive limitations. These are two forms of investigating two different notions of rationality. The first is an optimality-oriented axiomatic approach that abstracts the cognitive limitations of finite reasoners. The second is an ecological approach that considers those limitations. These projects have

Danilo Fraga Dantas

different methods and goals and should both be carried out, although the assumption of opinionation still calls for a justification from Bayesian epistemologists.

## References

- Babic, B. 2019. "A theory of epistemic risk". *Philosophy of Science* 86(3): 522–550.
- Baddeley, A., M. Eysenck, and M. Anderson. 2020. *Memory* (3rd ed.). Routledge.
- Bourne, R., and S. Parsons. 1998. "Propagating probabilities in System P". *Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference*: 440–445.
- Carr, J. 2015. "Epistemic expansions". *Res Philosophica* 92(2): 217–236.
- Cevolani, G., and R. Festa. 2021. "Approaching deterministic and probabilistic truth: A unified account". *Synthese* 199(3–4): 11465–11489.
- Cherniak, C. 1986. *Minimal Rationality*. MIT Press.
- Cooper, G. 1990. "The computational complexity of probabilistic inference using Bayesian belief networks". *Artificial Intelligence* 42: 393–405.
- Dantas, D. F. 2017. "No rationality through brute-force". *Philosophy South* 18(3): 195–200.
- . 2021. "How to (blind)spot the truth: An investigation on actual epistemic value". *Erkenntnis* 88: 693–720
- . 2022. "Epistemic sanity or why you shouldn't be opinionated or skeptical". *Episteme* 20(3): 647–666
- De Bona, G., and J. Staffel. 2017. "Graded incoherence for accuracy-firsters". *Philosophy of Science* 84(2): 189–213.
- . 2018. "Why be (approximately) coherent?". *Analysis* 78(3): 405–415.
- De Finetti, B. 1937. "La prévision: Ses lois logiques, ses sources subjectives". *Annales de l'institut Henri Poincaré* 7 (1), 1–68.
- Doets, K. 1994. *From Logic to Logic Programming*. MIT Press.
- Douven, I. 2013. "Inference to the best explanation, Dutch books, and inaccuracy minimization". *The Philosophical Quarterly* 63(252): 428–444.
- Dudai, Y. 1997. "How big is human memory, or on being just useful enough". *Learning & memory* 3(5): 341–365.
- Fitelson, B., and K. Easwaran. 2015. "Accuracy, coherence and evidence". In *Oxford studies in Epistemology, volume 5*, edited by Tamar Gendler and John Hawthorne (Eds.), 61–96. Oxford University Press.
- Friedman, J. 2013. "Suspended judgment". *Philosophical Studies* 162(2): 165–181.
- . 2015. "Why suspend judging?". *Noûs* 51(2): 302–326.

- Gao, J. 2021. "Self-deception and pragmatic encroachment: A dilemma for epistemic rationality". *Ratio* 34(1): 5–84.
- Garber, D. 1984. "Old evidence and logical omniscience in Bayesian confirmation theory". In *Testing Scientific Theories*, edited by John Earman, 99–131. University of Minnesota Press.
- Genin, K., and F. Huber. 2020. "Formal representations of belief". In *The Stanford encyclopedia of Philosophy (Winter 2020)*, edited by Edward Zalta. Metaphysics Research Lab, Stanford University.
- Gigerenzer, G. and W. Gaissmaier. 2011. "Heuristic decision making". *Annual Review of Psychology* 62(1): 451–482.
- Hájek, A. 2008. "Arguments for or against probabilism?". *The British Journal for the Philosophy of Science* 59(4): 793–819.
- Jeffrey, R. 1983. *The logic of decision*. University of Chicago Press.
- Joyce, J. 1998. "A nonpragmatic vindication of probabilism". *Philosophy of Science* 65(4): 575–603.
- Kemeny, J. 1955. "Fair bets and inductive probabilities". *Journal of Symbolic Logic* 20(3), 263–273.
- Kraus, S., D. Lehman, and M. Magidor. 1990. "Nonmonotonic reasoning, preferential models and cumulative logics". *Artificial Intelligence* 44(1-2): 167–207.
- Leitgeb, H. 2014. "The stability theory of belief". *The Philosophical Review*, 123(2): 131–171.
- Leitgeb, H., and R. Pettigrew. 2010. "An objective justification of Bayesianism II: The consequences of minimizing inaccuracy". *Philosophy of Science*, 77(2): 236–272.
- Lewis, P. J., and D. Fallis. 2019. "Accuracy, conditionalization, and probabilism". *Synthese* 198(5): 4017–4033.
- Littlejohn, C. 2015. "Who cares what you accurately believe?". *Philosophical Perspectives* 29(1): 217–248.
- Lord, E. 2020. Suspension of judgment, rationality's competition, and the reach of the epistemic. In *The Ethics of belief and beyond: Understanding mental normativity*, edited by S. Schmidt and G. Ernst, 126–145. Routledge.
- Pearl, J. 1986. "Fusion, propagation, and structuring in belief networks". *Artificial Intelligence*, 29: 241–288.
- Pettigrew, R. 2019. "Veritism, epistemic risk, and the swamping problem". *Australasian Journal of Philosophy* 97(4): 761–774.

Danilo Fraga Dantas

- Ramsey, F. 1926. "Truth and probability". In *The foundations of mathematics and other logical essays*, edited by R. Braithwaite, 156–198. McMaster University Archive.
- Russell, S., and P. Norvig. 2020. *Artificial intelligence: A modern approach (4th ed.)*. Pearson.
- Steele, K., and H. O. Stefánsson. 2020. "Decision theory". In *The Stanford encyclopedia of Philosophy (Winter 2020)*, edited by Edward Zalta. Metaphysics Research Lab, Stanford University.
- Stenning, K. and M. van Lambalgen. 2008. *Human reasoning and cognitive science*. MIT Press.
- Sturgeon, S. 2008. "Reason and the grain of belief". *Noûs*, 42(1): 139–165.
- Talbott, W. 1991. "Two principles of Bayesian Epistemology". *Philosophical Studies*, 62(2): 135–150.
- Talbott, W. 2016. "Bayesian Epistemology". In *The Stanford encyclopedia of Philosophy (Winter 2016)*, edited by Edward Zalta. Metaphysics Research Lab, Stanford University.
- Teller, P. 1973. "Conditionalization and observation". *Synthese*, 26(2): 218–258.
- Todd, P. M. and G. Gigerenzer. 2007. "Environments that make us smart: Ecological rationality". *Current Directions in Psychological Science*, 16(3): 167–171.
- Trpin, B., and M. Pellert. 2019. "Inference to the best explanation in uncertain evidential situations". *British Journal for the Philosophy of Science*, 70(4): 977–1001.
- Wedgwood, R. 2015. "Doxastic correctness". *Aristotelian Society Supplementary Volume* 87(1): 217–234.
- Wolpert, D., and W. Macready. 1997. "No free lunch theorems for optimization". *IEEE Transactions on Evolutionary Computation*, 1(1): 67–82.

# SUBJECTIVE RATIONALITY AND THE REASONING ARGUMENT

Erhan DEMIRCIOGLU

ABSTRACT: My main aim in this paper is to show that Kolodny's intriguing argument against wide-scopism – 'the Reasoning Argument' – fails. A proper evaluation of the Reasoning Argument requires drawing two significant distinctions, one between thin and thick rational transitions and the other between bare-bones wide-scopism (and narrow-scopism) and embellished wide-scopism (and narrow-scopism). The Reasoning Argument is intended by Kolodny both as an argument against bare-bones wide-scopism and as an argument against embellished wide-scopism. I argue that despite its formidable virtue of demonstrating the need for an account of thick subjective rationality, the Reasoning Argument works neither against bare-bones wide-scopism nor against embellished wide-scopism.

KEYWORDS: rational requirements, wide scope, narrow scope, subjective rationality, objective rationality

## 1. Introduction

Rationality, theoretical or practical, forbids incoherence among one's attitudes. It is irrational to be *akratic*: it is irrational of one not to intend to  $X$ , if one believes that one ought to  $X$ ; and, conversely, it is irrational of one to intend to  $X$ , if one believes that one ought not to  $X$ . It is irrational to be *means-end incoherent*: it is irrational of one not to intend to  $Y$ , if one intends to  $X$  and believes that  $Y$  is necessary for  $X$ . It is irrational to be *meta-incoherent*:<sup>1</sup> it is irrational of one to believe that  $p$  while also believing that it is irrational to believe that  $p$ .

The kind of rationality that depends on coherence relations among one's attitudes is sometimes called 'subjective' rationality – the contrast here being with 'objective' rationality, which is roughly a matter of being supported by one's reasons or evidence.<sup>2</sup> It is one thing to hold that one's attitudes are supported by one's reasons or evidence, and it is an at least nominally different thing to hold that

---

<sup>1</sup> The term 'meta-incoherence' is adopted from Huemer (2011).

<sup>2</sup> In this paper, I adopt Kolodny's (2005) terminology for the distinction under consideration between two kinds of rationality, as I will be discussing his views on the matter. However, it is worth noting that the terms preferred in recent years by the literature to express the distinction Kolodny is interested in are 'structural' and 'substantive' rationality, the former corresponding to the 'subjective' part and the latter 'objective.' See, for instance, Fogal & Worsnip (2021).

one's attitudes fit or match together well. The first claim concerns objective rationality, and the latter subjective rationality. Believing what one's reasons *in fact* support is the objectively rational thing to do, and believing what *one believes* one's reasons support is a subjectively rational thing to do.<sup>3</sup>

There is a major debate between wide-scopers and narrow-scopers about the nature of subjective rationality, which centers on the logical form of its requirements. Wide-scopers hold that the 'rationality requires' operator always takes wide scope over an entire conditional and that a requirement of subjective rationality always takes the following form: rationality requires one (to *Y*, if one *Xs*).<sup>4</sup> This is equivalent to claiming that rationality requires one *either* not to *X* *or* to *Y*: no particular attitude is required. So, for instance, a wide-scoper for akratic constraints on rationality holds that if I find myself believing that I ought to quit smoking while lacking the intention to do so, I am required, subjectively speaking, either to abandon the belief in question or form the relevant intention. Narrow-scopers, on the other hand, hold that the 'rationality requires' operator sometimes takes narrow scope over the consequent of the relevant conditional and that a requirement of subjective rationality sometimes takes the following form: if one *Xs*, then rationality requires one to *Y*.<sup>5</sup> This means that, according to narrow-scopers, rationality sometimes requires one to take a particular attitude. So, for instance, a narrow-scoper for akratic constraints on rationality might hold that if I find myself believing that I ought to quit smoking while lacking the intention to do so, I am required, subjectively speaking, to form the relevant intention: abandoning the belief is not the correct subjective response to the incoherent combination of attitudes I happen to have, despite the fact that by abandoning the belief, I end up with a *coherent* state.

---

<sup>3</sup> It is important to note that the terms 'structural' and 'substantive' capture the point of the relevant distinction better than the terms 'subjective' and 'objective' do. The point of the relevant distinction is that rationality as it pertains to the *coherence relations* between one's attitudes is to be distinguished from rationality as it pertains to the support relations between one's attitudes and reasons. Contrary to what the term 'subjective rationality' suggests, the 'domain' of coherence as such is not restricted to the relations that hold between a subject's attitudes and her 'perspective' on (or her beliefs about) whether those attitudes are supported by reasons. It is, for instance, incoherent to believe that *p* is true and also to believe that *p* is false, and the incoherence here has nothing to do with the subject's 'perspective' on whether those two beliefs are supported by reasons.

<sup>4</sup> For recent defenses of wide-scopism, see, for instance, Broome (1999; 2007), Brunero (2010; 2012), Wallace (2001), and Way (2011). For earlier presentations and defenses of this view, see Darwall (1983), Greenspan (1975), and Hill (1973).

<sup>5</sup> For a defense of narrow-scopism, see, for instance, Bedke (2009), Kolodny (2005; 2007), Lord (2014), Schroeder (2004; 2009),



The debate between wide-scopers and narrow-scopers concerns the scope of the *process*-requirements (as opposed to *state*-requirements) of subjective rationality. State-requirements “simply ban states in which one has conflicting attitudes,” while process-requirements say “how, going forward, one is to form, retain, or revise one’s attitudes so as to avoid or escape such conflict-states” (Kolodny 2005, 517).<sup>6</sup> State-requirements are in the business of identifying conflict-states, and process requirements are in the business of identifying ways to avoid or resolve those conflict-states. We can judge whether a subject is rational or not by virtue of the attitudes she has at a given time, and we can also judge whether she is rational or not in virtue of how she transitions from one state to another over time. The former judgment is based on state-requirements, and the latter process-requirements.

State-requirements have, as Kolodny notes (2005, 540), wide scope because a ban on the combination of having *X* and *Y* can be satisfied either by not having *X* or by not having *Y*: there is no ‘asymmetry’ between these two ways of satisfying the ban in question. If I don’t have *X* but have *Y*, and you don’t have *Y* but have *X*, then you and I both satisfy the state-requirement ‘Don’t be in *X* & *Y*’. So, the debate between the narrow-scopers and the wide-scopers can only be about the scope of process-requirements. Thus taken, the narrow-scopers hold that there are some conflict-states that can only be rationally escaped in one way, by revising a particular conflicting state, whereas the wide-scopers hold that there are no such conflict-states, that all conflict-states can be rationally escaped in (at least) *two* ways, by revising one or the other of the conflicting states.

My main aim in this paper is to show that Kolodny’s intriguing argument against wide-scopism, which I call ‘the Reasoning Argument’, fails. Though I hold the Reasoning Argument in high regard, I also find it extremely dense and believe that it can appreciably benefit from some considerable unpacking. A proper evaluation of the Reasoning Argument requires drawing two significant distinctions, one between thin and thick rational transitions and the other between bare-bones wide-scopism (and narrow-scopism) and embellished wide-scopism (and narrow-scopism). The Reasoning Argument is intended by Kolodny both as an argument against bare-bones wide-scopism and as an argument against embellished wide-scopism. I argue that despite its formidable virtue of demonstrating the need for an account of thick subjective rationality, the Reasoning Argument works neither against bare-bones wide-scopism nor against embellished wide-scopism.

---

<sup>6</sup> All page references below are to Kolodny (2005), unless otherwise noted.

## 2. The Reasoning Argument

One of Kolodny's claims that plays a critical role in the Reasoning Argument concerns what it is for a transition to be rational. Process-requirements are concerned with rational transitions from some conflict-states to some other non-conflict-states over time, and Kolodny contends that some transitions from a conflict-state to a non-conflict-state are not rational. He writes:

A bolt of lightning might jolt me out of a state in which I have two inconsistent beliefs and into a state in which I lack one or both of them. Although this process might be said to take me from an irrational state to a rational one, it would not, itself, be rational. (2005, 517)

According to Kolodny, then, a rational transition is not simply a transition from an incoherent state to a coherent state: *the way in which* the transition is made from an incoherent state to a coherent one is a significant factor that determines whether the transition is rational or not. Coming to resolve a conflict-state as a result of being hit by lightning is not a rational transition. And, since process-requirements specify the conditions under which a transition from one state to another is rational, Kolodny holds, their correct formulations must respect the idea that how the transition is made matters to its rationality.

There are two things it pays to be careful here. One pertains to the notion of rationality Kolodny has in mind when he claims that some transitions from an incoherent state to a coherent one are not rational. As observed above, rationality comes in two kinds, objective and subjective; so, correspondingly, we can speak of objectively rational transitions and subjectively rational transitions. There are surely some transitions from an incoherent state to a coherent one that are not objectively rational. For instance, the transition from an incoherent state  $X \& Y$  to a coherent state  $X$  is not objectively rational, if  $X$  is not supported by reasons but  $Y$  is. However, Kolodny's point is obviously not about objectively rational but subjectively rational transitions, *viz.* that some transitions from an incoherent state to a coherent one are not *subjectively* rational, a thesis which I will hereafter call 'SRT'.

In order for SRT to be true, there must be something that is required for a transition to be subjectively rational (and not necessarily objectively rational) over and above the fact that it results with a resolution of a conflict-state. The objective rationality of a transition plausibly requires adequate sensitivity to epistemic support relations; and, accordingly, the subjective rationality of a transition might be plausibly conceived as demanding adequate sensitivity to coherence relations. However, if a transition in a given case results with a resolution of a conflict-state, is not it thereby adequately sensitive to the coherence relations operative in that

case? What more can it be plausibly required for it to be subjectively rational? The answers to these questions – which I will, *inter alia*, address in what follows – are not obvious; so, it is not obvious at this point that SRT is true.

The other thing I wish to say pertains to the risk of question-begging: SRT can be adequately deployed in an argument against wide-scopism only if it does not beg question against wide-scopism. Now, there is a clear sense in which wide-scopism is committed to *rejecting* SRT: there are (at least) two ways to escape an incoherent state, by revising either one or the other of the two conflicting attitudes; and, wide-scopism says that both are subjectively rationally acceptable, which amounts to saying that (there is a sense in which) all transitions from an incoherent state to a coherent one are subjectively rational. Furthermore, there is a clear sense in which narrow-scopism is committed to *accepting* SRT: there are two ways to escape an incoherent state; and, narrow-scopism says that there are some cases in which only one of those two ways is subjectively rationally acceptable, which amounts to saying that (there is a sense in which) not all transitions from an incoherent state to a coherent one are subjectively rational.<sup>7</sup> In order for SRT to play a dialectically effective role in an argument against wide-scopism/for narrow-scopism, what it means to be asserting should not prejudge the outcome of the debate between these two positions in one way or another. So, how exactly should we understand SRT?

Kolodny's defense of SRT proceeds by appealing to certain cases in which the transition from an incoherent state to a coherent one occurs as a result of what we might call 'brute' or 'mere' causes such as being hit by the lightning or given an electric shock. This appeal can be deployed in an attempt to specify what Kolodny means to be asserting by SRT, which might plausibly go as follows:

**CSRT** Merely causal transitions from an incoherent to a coherent state are not subjectively rational.

CSRT does not prejudge the debate between wide-scopism and narrow-scopism. The 'bare-bones' versions of these two positions, the versions which I have been taking for granted so far, are equally *incompatible* with CSRT. Neither bare-bones wide-scopism nor bare-bones narrow-scopism excludes the possibility that the rational transitions sanctioned by their requirements are merely causal. Consider the following requirements:

(WS-BB) If a subject believes that she ought to *X* but does not intend to *X*, then rationality requires her either to form the intention to *X* or to drop the belief that

---

<sup>7</sup> Henceforth, I will, for convenience's sake, sometimes omit the qualifier 'subjective' and I will always mean subjective rationality by 'rationality' unless otherwise noted.

Erhan Demircioglu

she ought to  $X$ .

(NS-BB) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her to form the intention to  $X$ .<sup>8</sup>

WS-BB is a bare-bones wide scope requirement, and NS-BB is a bare-bones narrow scope requirement. Obviously, there is no mention in either of these two formulations of how the transition from the conflict-state to a non-conflict-state must occur: they only tell us that the transition must *somehow* occur. This means that both NS-BB and WS-BB allow the possibility that the rational transition from an akratic state to forming the relevant intention while keeping the relevant 'ought-belief' might occur as a result of an electric shock.

Furthermore, both bare-bones wide-scopism and bare-bones narrow-scopism can be straightforwardly revised in a way to accommodate CSRT. A suitably revised wide scope requirement might go as follows: if one finds oneself in an incoherent state  $X$  &  $Y$ , then rationality requires that either  $X$  or  $Y$  is dropped *in a way that is not merely causal*. Similar remarks apply to the bare-bones narrow scope version of this schema. Hence, we get the following theses for akrasia:

(WS-NC) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her either to form the intention to  $X$  in a way that is not merely causal or to drop the belief that she ought to  $X$  in a way that is not merely causal.

(NS-NC) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her to form the intention to  $X$  in a way that is not merely causal.

CSRT can thus be accommodated, equally directly, by wide-scopism and narrow-scopism.

There is a point that is worth emphasizing here. In order for the accommodation of CSRT by wide-scopism to be genuine but not a mere verbal façade, it should be possible to take either of the routes identified by a given wide scope requirement of resolving a certain conflict. Take for instance WS-NC, and suppose that it is not possible for a given subject to drop the belief that she ought to  $X$  in a way that is not merely causal. If this is so, then what rationality in effect requires of an akratic subject is to form the intention to  $X$  in a way that is not

---

<sup>8</sup> Kolodny's discussion in (2005) focuses on the 'reason' (rather than the 'ought') versions of these requirements. So, for instance, his version of (NS-BB) goes like this: if one believes that one has conclusive *reason* to  $X$ , then rationality requires one to intend to  $X$  (p. 528). In (2007, 3), however, the focus is on the 'ought' versions. Nothing much hangs on this distinction for the purposes of this paper, and for convenience's sake, I prefer the 'ought' versions presented above.

merely causal, which means that WS-NC collapses into NS-NC and narrow-scopism wins the day. However, there is no reason at this point for wide-scopism to worry about this “collapsing threat” because we have not yet been given any reason to think that it is not possible to take one of the routes identified by WS-NC (or by other wide scope requirements that accommodate CSRT).

Having argued that CSRT does not appear to prejudice the debate between wide-scopism and narrow-scopism, let me briefly note that it has a counterpart thesis regarding the objective rationality of a transition:

**CORT** Merely causal transitions from a state unsupported by one’s reasons to a supported one are not objectively rational.

Consider the following case. I believe that  $p$ , while my evidence supports not- $p$ . Suppose that I come to drop the belief that  $p$  and form the belief that not- $p$  as a result of an electric shock. My transition in this case is merely causal, and it appears that there is a clear sense in which this transition is not objectively rational, despite the fact that it ends up with my coming to form an attitude supported by my evidence (more on this later).

It is fair to proceed, then, by granting CSRT. What is it, according to Kolodny, that explains why CSRT is true? What is it about a merely causal transition that disqualifies it from being subjectively rational? Suppose that I am in a conflict-state  $X \& Y$ , and then come to drop  $Y$ , thereby resolving the conflict, as a result of being hit by lightning. This is not, CSRT entails, a subjectively rational transition; and what is missing, Kolodny notes (p. 520), seems to be *reasoning*: the transition from  $X \& Y$  to  $X$  would be a subjectively rational transition if, for instance, it resulted from the fact that, having recognized the conflict between  $X$  and  $Y$ , I *reasoned* to revising  $Y$ . So, it seems that the way in which a subjectively rational transition is made has to do with the fact that the subject goes through a process of reasoning. On Kolodny’s account, a given rationality process-requirement can be adequate only if the transitions it requires of a subject to resolve a given conflict can be made via reasoning.

So, we arrive at the following thesis:

**RSRT** Transitions from an incoherent to a coherent state that are not based on reasoning are not subjectively rational.

There are four quick points I want to make here. Firstly, I take it that RSRT is purported to provide an explanation of why CSRT is true: CSRT is true, the intended explanation goes, because RSRT is true. How does RSRT explain that CSRT is true? One option is to maintain that RSRT explains CSRT because what it is for a transition to be merely causal is simply for it not to be reasoning-based (i.e.,

because a transition is merely causal *if, and only if*, it is not reasoning-based). However, this equivalence-thesis is unnecessarily strong for present purposes.<sup>9</sup> Another, weaker option is to maintain that RSRT explains CSRT because *if* a transition is merely causal, then it is not reasoning-based (or, equivalently, there are no merely causal but reasoning-based transitions). If this weaker sufficiency-thesis is true, then it follows from RSRT and the fact that a given transition is merely causal, that that transition is not subjectively rational, which is also the verdict CSRT delivers about that transition.

Secondly, it is plausible that if a transition is merely causal, then it is not reasoning-based, so the move from CSRT to RSRT appears to be acceptable. It seems clear that there is a substantive difference between reasoning-based transitions and merely causal transitions. Compare the following three transitions from  $X \& Y$  to  $X$ . One is based on reasoning, another is caused by a head injury, and the final one is caused by a bolt of lightning. Despite the fact that each of these transitions differs at a certain fine-grained level from the others, a natural classificatory strategy would be to treat the one based on reasoning as substantially different from the other two. The naturalness of this classification plausibly reflects the fact that if a transition is merely causal, then it is not reasoning-based.<sup>10</sup>

Thirdly, the plausibility of the move from CSRT to RSRT does *not* require that reasoning not be a causal process. Reasoning might be causal without being 'merely' causal (whatever that exactly comes to), and CSRT claims only that merely causal transitions cannot be subjectively rational and does not thereby exclude the possibility that some causal transitions are subjectively rational.

---

<sup>9</sup> Additionally, the equivalence thesis at hand appears to be independently implausible. It seems that there might be some transitions that are neither reasoning-based nor merely causal. If my wish that my son will be a good person causes my belief that he will be a good person, then it seems that the transition from my wish to my belief is not reasoning but it is not merely causal either. This transition at least appears to be substantially different from the sort of merely causal transitions (e.g., coming to form an attitude as a result of an electric shock) that Kolodny points out we need to distinguish from rational ones.

<sup>10</sup> There might be different ways to account for the fact that a merely causal transition is not reasoning-based and our classificatory intuitions regarding the cases presented. A prominent one, an agency-oriented explanation, goes roughly as follows: reasoning-based transitions is something that we 'do' whereas merely causal transitions is something that 'merely happens to us', which grounds our attributions of praise and blame to the former ones but not to the latter. It is because reasoning-based transitions are something we do, the agency-oriented explanation goes, that a merely causal transition, which cannot ground our attributions of praise and blame, is not reasoning-based. For the purposes of this paper, I remain non-committal about what it is that ultimately explains that a merely causal transition is not reasoning-based.

Fourthly, there appear to be cases for which CSRT and RSRT fail to deliver any clear verdict about the subjective rationality of a given transition. For instance, consider the following scenario. Suppose that having recognized the conflict between  $X$  and  $Y$ , I have reasoned to the conclusion that  $Y$  is to be dropped. However, suppose further that I find myself unable to drop  $Y$  as my psychological attachment to it proves to be stronger than my willingness to drop it. Under these circumstances, let us further suppose, I deliberately take an electric shock, knowing that its effect will be dropping  $Y$ . Now, what can we say of the *entire* process that ends up with dropping  $Y$ ? Is it merely causal or reasoning-based? The process taken in its entirety has an initial reasoning-based stage and a subsequent merely causal stage, and it seems that we cannot say either that it is, as a whole, merely causal *simpliciter* or that it is, as a whole, reasoning-based *simpliciter*. If this is so, then we cannot get any clear guidance about the subjective rationality of this process either from CSRT or from RSRT. However, as I don't think such cases as these cast any serious doubt on the plausibility of CSRT or of RSRT (but I think they call for an account of the individuating conditions of 'a transition'), I will not base my attack to the Reasoning Argument on the possibility of those cases.

Having made these clarificatory remarks about the relationship between CSRT and RSRT, let me raise the following question: can RSRT be deployed in an attempt to arbitrate between wide-scopism and narrow-scopism? Consider the following wide scope and narrow scope requirements for akrasia that accommodate RSRT:

(WS-NR) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her either to form the intention to  $X$  in a reasoning-based way or to drop the belief that she ought to  $X$  in a reasoning-based way.

(NS-NR) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her to form the intention to  $X$  in a reasoning-based way.

The collapsing threat that we have pointed out arises in regard to wide scope requirements that accommodate CSRT rears its head once again but now, one might plausibly think, *much more glaringly*. Suppose that it is not possible for an akratic subject to reason to dropping her ought-belief. If this is so, then WS-NR collapses into NS-NR and narrow-scopism wins the day. Furthermore, it might as well not be possible for an akratic subject to reason to dropping the ought-belief because it seems plausible that there are some constraints on a reasoning process, some of which might not be satisfied by the subject's transition to dropping her ought-belief. Reasoning is an evidently demanding cognitive process, and it is this evident demanding character that renders the collapsing threat for wide-scopism much more glaring than before.

Kolodny's take on the collapsing threat for wide-scopism is formulated in terms of what he calls 'the Reasoning Test' (pp. 520-521), which goes as follows:

**The Reasoning Test:** The process-requirement governing the conflict between  $X$  and  $Y$  is wide scope only if (i) it is possible to reason from the content of  $X$  to dropping  $Y$  and (ii) it is possible to reason from the content of  $Y$  to dropping  $X$ .

According to the Reasoning Test, if one cannot reason, for instance, from the content of  $X$  to dropping  $Y$  but can only reason from the content of  $Y$  to dropping  $X$ , then the process-requirement governing the conflict between  $X$  and  $Y$  is not wide scope but can only be narrow scope. In other words, if either (i) or (ii) is false, then the wide scope version of the process requirement in question collapses into the relevant narrow scope version, and narrow-scopism wins.

Here is how the Reasoning Argument goes, focusing on WS-NR. Consider the conflict-state of believing that she ought to  $X$  while not intending to  $X$ , which the wide-scooper thinks is governed by WS-NR. In order for WS-NR to pass the Reasoning Test, it must be possible to reason from the content of not intending to  $X$  to dropping the ought-belief and also possible to reason the other way around. However, it is not possible, Kolodny argues (p. 528), to reason from the content of not intending to  $X$  to dropping the ought-belief; and this is because not intending to  $X$  (or lacking an attitude in general) has no content. So, WS-NR collapses into NS-NR. However, it is possible, Kolodny notes (p. 527), to reason from the belief that there is conclusive reason to  $X$  to forming the intention to  $X$ . So, NS-NR is the requirement governing the given conflict.

The case for narrow-scopism can be reinforced, Kolodny thinks, by considering the 'converse'-akratic wide scope and narrow scope requirements:

(WS-CNR) If a subject believes that she ought not to  $X$  but intends to  $X$ , then rationality requires her either to drop the intention to  $X$  in a reasoning-based way or to drop the belief that she ought not to  $X$  in a reasoning-based way.

(NS-CNR) If a subject believes that she ought not to  $X$  but intends to  $X$ , then rationality requires her to drop the intention to  $X$  in a reasoning-based way.

Here is how the Reasoning Argument goes, focusing on WS-CNR. Consider the conflict-state of believing that one ought not to  $X$  while intending to  $X$ , which the wide-scooper thinks is governed by WS-CNR. If WS-CNR has wide scope, then it must pass the Reasoning Test: it must be possible to reason from the content of the intention to  $X$  to dropping the belief that she ought not to  $X$  and also possible to reason the other way around. However, it is not possible, Kolodny argues, to reason from the content of the intention to  $X$  to dropping the ought-belief in question because reasoning is, in his words, not 'upstream:' "it is not reasoning to



cling to what one judges to be an unfounded intention and to support it by revising one's belief about one's reasons" (p. 529), it is some other process, "such as self-deception or wishful thinking" (p. 530). So, WS-CNR collapses into NS-CNR. However, it is possible, Kolodny notes, to reason from the belief that one ought not to  $X$  to dropping the intention to  $X$ . So, NS-CNR is the requirement governing the given conflict.<sup>11</sup>

The Reasoning Argument is ingeniously crafted. It starts with a thesis, CSRT, that is apparently neutral between narrow-scopism and wide-scopism. And, from CSRT, it convincingly moves to RSRT. Surprisingly, despite the fact that CSRT is apparently neutral and the move from CSRT to RSRT is evidently plausible, Kolodny shows that RSRT can be deployed in an argument against wide-scopism. This is, Kolodny goes on to argue, because there appears to be some substantive constraints on what it takes to reason, constraints that can be uncovered with some analytical ingenuity. The guiding insight is that reasoning is not just any mental transition from one attitude to another. More specifically, in order a process to be a process of reasoning, there must be something (namely, *content*) to reason from (in a slogan form: all reasoning is content-processing) and that something must be processed in a certain way (in a slogan form: upstream content-processing is not reasoning). The Reasoning Argument attempts to show that once these constraints are adequately appreciated, it can be seen that there are some conflict-cases in which the wide scope requirements that one might think govern them collapse into their narrow scope versions.

### 3. Bare-Bones Wide-Scopism Goes Unscathed

I will argue in this section that the starting move of the Reasoning Argument – namely, CSRT – can be resisted by wide-scopism. The wide-scooper can plausibly dig her heels in and stick with *bare-bones* wide-scopism, a position incompatible as I have maintained in the previous section with CSRT. If bare-bones wide-scopism does not need to be revised to accommodate CSRT, then the Reasoning Argument does not get off the ground.

Let us start with granting that there is a clear sense in which CSRT is true, that there is a sense of 'subjectively rational' in which a subjectively rational

---

<sup>11</sup> It must be clear that the argument from 'the downstream character' of reasoning also applies to WS-NR because if it were possible to reason, *per impossibile* according to Kolodny, from the content of *not* intending to  $X$  to revising the belief that one has conclusive reason to  $X$ , that reasoning would be 'upstream,' which is itself deemed impossible by that argument. So, WS-NR collapses, on Kolodny's account, into NS-NR not only because it is, implausibly, entitled to reasoning from lack of content but also because it is, implausibly, entitled to upstream reasoning.

transition requires more than moving from an incoherent state to a coherent one: *how* the transition is made, whether it is merely causal or not, matters to the rationality of the transition in this sense of the term. So, on this sense of a rational transition, a transition caused by a bolt of lightning is not subjectively rational even if it is a transition from an incoherent to a coherent state. Let us call the sense in which merely causal transitions are not subjectively rational, the sense in which CSRT is true, the *thick* sense of a subjectively rational transition.

However, there is another, less demanding, and still perfectly intelligible sense of 'subjectively rational' in which it suffices for a transition to be subjectively rational that it is from an incoherent state to a coherent one. If, for instance, I believe that I ought to *X* but do not intend to *X* at a certain time, and then, at a later time, I come to drop the ought-belief or come to form the relevant intention, then I have surely made a subjectively *rational* progress: my attitudinal system has now become more coherent and thereby more rational than before. On this conception of a subjectively rational transition, a given transition is subjectively rational just in case it is a process as a result of which one's attitudes come to comply with the requirements of coherence. After forming the intention to *X*, for instance, I come to comply with the relevant coherence requirement I am under and therefore the transition that I go through ending with compliance is rational in *this* sense. Let me call the sense in which all transitions from an incoherent to a coherent state are rational, the sense in which CSRT is false, the *thin* sense of a rational transition.

Consider my friend Mary, whose attitudinal system is imbued with many inconsistencies at a certain time, and suppose that being frustrated with those inconsistencies, I decide not to see her for a few months. Months pass, and we meet again. Now, I realize, to my pleasure, that Mary has become much more coherent than she was before: those inconsistencies that were once so frustrating to me are now removed from her attitudinal system. I don't know *how* she has managed to pull that off, but I am glad that it is done, one way or another. Surely, I can now correctly say that Mary has made a subjectively *rational* progress, despite the fact that I don't know how – perhaps, a benevolent demon intervened and made the relevant revisions, or perhaps, she carefully reflected on those inconsistencies and dedicated some mental effort to eradicate them, or perhaps, she was hit by lightning, which caused the necessary changes in her brain states. There is a clear and intelligible sense of a subjectively rational progress in which *how* Mary has managed to make the progress does not matter to the accuracy of my judgment that Mary has made a rational progress – and this is the *thin* sense of a subjectively rational progress.

On the thin sense of a subjectively rational transition, all transitions from an incoherent to a coherent state are subjectively rational, and CSRT is false. It is instructive here to compare CSRT with its objective counterpart introduced above:

**CORT** Merely causal transitions from a state unsupported by one's reasons to a supported one are not objectively rational.

When I move from the belief that  $p$ , unsupported by my evidence, to the belief that not- $p$ , supported by my evidence, there is a clear sense in which I make an objectively rational progress, irrespective of whether the progress is due to a mere cause or not. It is on the thin sense of an objectively rational transition that my transition is objectively rational, and it is on this sense that CORT is false.

The distinction I draw between the thin and the thick senses of an objectively rational transition can be plausibly traced back to a distinction commonly drawn between two types of objective (epistemic) justification, i.e., “propositional” versus “doxastic” justification.<sup>12</sup> If I have good evidence that supports a belief, I am propositionally (objectively) justified to have that belief, irrespective of whether I have it or not and *irrespective of whether I have formed it in a particular manner or not*. However, if I come to form that belief in a deficient manner (for instance, in a merely causal way), then I am not doxastically justified in having that belief (despite the fact that I have that belief and I have evidence that supports that belief). So, the objective rationality, in the doxastic sense, of a transition is sensitive to the way in which a given belief is formed. However, we can also intelligibly speak of the objective rationality, in the propositional sense, of a transition. If I have good evidence that supports a belief at a certain time and if I come to have that belief at a later time, then the transition that I thereby make is objectively rational, in the propositional sense, irrespective of the particular manner in which I come to have that belief. What *solely* matters to the objective rationality, in the propositional sense, of a transition is that it results with forming attitudes that are supported by one's evidence – and this is the thin sense of an objectively rational transition.

Now, a parallel distinction can be drawn with respect to subjectively rational transitions, in this case between propositional and what one might call ‘attitudinal’<sup>13</sup> rationality. We can say that if a certain attitude coheres with my

---

<sup>12</sup> For a book-length discussion on the nature of the distinction between propositional and doxastic justification, see Silva & Oliveira (2022).

<sup>13</sup> Since doxastic rationality, by definition, pertains only to the rationality of doxastic attitudes such as beliefs, and since subjective rationality pertains to the coherence of all sorts of attitudes including intentions as well as beliefs, the term we need here to cover the opposing class must be broader in its application than ‘doxastic.’

current attitudes, then it is propositionally (subjectively) rational for me to have that attitude, irrespective of whether I have it or not and *irrespective of whether I come to have it in a particular manner or not*. However, if I come to have that attitude in a deficient manner (for instance, in a merely causal way), then I am not attitudinally (subjectively) rational in having that attitude (despite the fact that I have that attitude and that attitude coheres with the rest of my attitudes). So, the subjective rationality, in the attitudinal sense, of a transition is sensitive to the way in which a given belief is formed. However, we can also intelligibly speak of the subjective rationality, in the propositional sense, of a transition. If an attitude coheres with the rest of my attitudes at a certain time, and if I come to have that attitude at a later time, then the transition that I thereby make is subjectively rational, in the propositional sense, irrespective of the particular manner in which I come to have that attitude. What *solely* matters to the subjective rationality, in the propositional sense, of a transition is that it results with forming attitudes that cohere with the rest of my attitudes – and this is the thin sense of a subjectively rational transition.

Given the distinction between the two senses of a subjectively rational transition – the ‘thick’ (or ‘attitudinal’) sense Kolodny endorses and the ‘thin’ (or ‘propositional’) sense I have articulated – we are in a position to see that it does not follow, contrary to Kolodny’s claim, from the fact that process-requirements are in the business of identifying subjectively rational transitions, that their correct formulations must respect the idea that how the transition is made matters to its subjective rationality. If process-requirements are in the business of identifying subjectively rational transitions in the *thin* sense, then their correct formulations might be entirely ‘oblivious’ to the way in which those transitions are made – or in other words, bare-bones wide-scopism might be true.

The fact that there is an intelligible sense of a rational transition that is compatible with wide-scopism – *viz.* the thin sense – suffices to undermine the Reasoning Argument as an argument against bare-bones wide-scopism. The Reasoning Argument assumes that there is only *one* intelligible sense of a rational transition and it is the thick sense, purported to be captured by CSRT. If this assumption is true, then bare-bones wide-scopism is not true because bone-bones wide scope requirements are designed to be oblivious to the way in which a transition from an incoherent state to a coherent one is made. However, the assumption is not true as there is also the thin sense of a rational transition, and the fact that there is *an* intelligible sense of a rational transition that is incompatible with bare-bones wide-scopism – *viz.* the thick sense – does not show that bare-bones wide-scopism is false. The upshot is that, in response to the Reasoning

Argument, wide-scopism can maintain its original bare-bones position, resting its case on the thin sense of a rational transition.

According to Kolodny, the explanation, at least in part, of why many (including, as he openly states, his former self) have failed to see that some rational requirements have narrow scope is that they focus exclusively on state-requirements, which after all have wide scope, and do not recognize the category of process-requirements. He writes: “It is only once we turn from state-requirements to process-requirements – only once we shift our focus from the rationality of synchronic states to the rationality of diachronic transitions among them – that we come to see that some rational requirements have narrow scope” (pp. 454-455). However, it must now be clear that this explanation will not do: turning from state-requirements to process-requirements explains the failure to recognize that some rational requirements have narrow-scope *only if* the thick sense of a rational transition is taken for granted as the only sense that can be legitimately appealed to in the course of formulating process-requirements. However, the thick sense of a rational transition is, as I have argued, not the only relevant sense of a rational transition.

Not only wide-scopers can retain their original bare-bones position by an appeal to the thin sense of a subjectively rational transition, but they *actually* take that sense for granted, which can be effectively illustrated by Broome’s (2007) response to Kolodny (2005).<sup>14</sup> Consider WS-BB again:

(WS-BB) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her either to form the intention to  $X$  or to drop the belief that she ought to  $X$ .

Kolodny’s objection that WS-BB does not capture the relevant rationality requirement one is under is based on the observation that there are not two processes of *reasoning* by means of which one can bring oneself to satisfy it (and this in turn is because one cannot reason, as we have noted above Kolodny argues,

---

<sup>14</sup> In his paper, Broome complains, rightly in my opinion, that he cannot “get [any] guidance from” Kolodny’s (2005) paper about “how to understand [a given requirement] as a requirement on processes” because, he notes, “none of the formulae for requirements of rationality set out in [that] paper mention processes; they all mention states only” (2005, 366). In his response, Kolodny accepts Broome’s point and goes on to provide “more explicit formulations” (2007, 3) of the process-requirements that he has in mind. Notwithstanding the soundness of Broome’s complaint about the unclarity of the content of process-requirements in Kolodny (2005), I believe that his ensuing defense of wide-scopism against the Reasoning Argument succeeds in getting across a valuable lesson about the (bare-bones) wide-scopers’ take on process-requirements.

from the absence of intention to dropping the belief). In his response, Broome simply grants Kolodny's observation, noting that "you can reason in one direction but not the other" (2007, 366), but he goes on to warn against equivocating on 'ways of satisfying [WS-BB].' Broome emphasizes that "there are two ways of satisfying [WS-BB] in the sense in which there are two ways in which a material conditional can be true: by its antecedent's being false or by its consequent's being true" (2007, 366). Broome's point against Kolodny is that there simply being two ways of satisfying WS-BB on account of the fact that it is a requirement to make a conditional true (and not whether there are two processes of reasoning by means of which one can bring oneself to satisfy it) is what is relevant to an assessment of WS-BB as a wide-scope requirement. And, Broome's emphasis on there being two ways to satisfy WS-BB can be plausibly read as involving a commitment to the claim that how the transition is made from an incoherent to a coherent state does *not* matter to the subjective rationality of the transition, and this claim is true on the thin sense of a subjectively rational transition.

I have argued in this section that the Reasoning Argument fails as an argument against bare-bones wide-scopism because it fails to acknowledge that there is a thin sense of a rational transition which is taken for granted by bare-bones wide-scopers. I anticipate three objections. One is that there is no such thing as the thin sense of a rational transition. Another one acknowledges the thin sense of a rational transition but holds that it is not relevant to subjective rationality. And, the last one acknowledges that the thin sense is relevant to subjective rationality but maintains that the process-requirements of subjective rationality govern not only rational transitions in the thin sense but also rational transitions in the thick sense. In what follows, I will not address the first two objections since they are, in my opinion, evidently mistaken: we can intelligibly speak of the rationality of a given transition from an incoherent to a coherent state without thereby committing ourselves to the character of the particular way in which the transition is achieved; and, the locus of that sort of rationality is clearly subjective rationality as it is about maintaining coherence. The third objection, on the other hand, raises a host of interesting and unexpected issues, which I will discuss in the next section.

#### **4. Wide-Scopism Can Account for Thick Subjective Rationality**

There being the thin (or propositional) sense of a rational transition undermines the Reasoning Argument *qua* an argument against bare-bones wide-scopism, the original wide-scooper position. Bare-bones wide-scopism goes unscathed, and this is because the (thin/propositional) sense of a rational transition it takes for granted is

not the (thick/attitudinal) sense of a rational transition the Reasoning Argument rests on, which means that the argument owes whatever apparent force it might have to equivocating on 'rational.' Bare-bones wide-scopism is an account of *thin* subjective rationality, and pointing at its expectable failure to account for *thick* subjective rationality cuts no ice against it.

The objection I wish to consider now acknowledges that the Reasoning Argument fails as an argument against bare-bones wide-scopism, but it goes on to claim that the argument points towards the need to account for thick subjective rationality, which bare-bones wide-scopism is not fit to meet. And, the Reasoning Argument shows, the objection goes, that what one might call *embellished* narrow-scopism (that is, the sort of narrow-scopism that aims to account for thick subjective rationality) is to be preferred over what one might call *embellished* wide-scopism (that is, the sort of wide-scopism that aims to account for thick subjective rationality). The following requirements, which I have introduced in Section 2, are embellished requirements:

(WS-NR) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her either to form the intention to  $X$  in a reasoning-based way or to drop the belief that she ought to  $X$  in a reasoning-based way.

(WS-CNR) If a subject believes that she ought not to  $X$  but intends to  $X$ , then rationality requires her either to drop the belief that she ought not to  $X$  in a reasoning-based way or to drop the intention to  $X$  in a reasoning-based way.

(NS-NR) If a subject believes that she ought to  $X$  but does not intend to  $X$ , then rationality requires her to form the intention to  $X$  in a reasoning-based way.

(NS-CNR) If a subject believes that she ought not to  $X$  but intends to  $X$ , then rationality requires her to drop the intention to  $X$  in a reasoning-based way.

Despite its failure as an argument against bare-bones wide-scopism, the Reasoning Argument shows, the objection goes, that embellished wide scope requirements, WS-NR and WS-CNR, collapse into embellished narrow scope requirements, NS-NR and NS-CNR, respectively.

The wide-scooper might be tempted to reply to this objection by simply turning a blind eye to the need effectively emphasized by Kolodny to have embellished versions of bare-bones requirements.<sup>15</sup> I think this temptation should be resisted. I take Kolodny to have persuasively demonstrated that there is such a thing as the thick (or attitudinal) sense of subjective rationality (as well as the thin (or propositional) sense of subjective rationality); and sticking with the thin (or propositional) sense of subjective rationality, which is taken for granted by bare-

---

<sup>15</sup> This is, so far as I can see, the line taken by Way (2011, see especially 237-238).

bones wide-scopism, does not make the need for an account of the thick (or attitudinal) sense of subjectively rationality disappear.

So, how does the Reasoning Argument fare as an argument against *embellished* wide-scopism? My answer is “badly.” Neither WS-NR collapses into NS-NR, nor WS-NCR into NS-NCR. Contra what the Reasoning Argument aims to establish, it is possible to drop the belief that one ought to  $X$  in a reasoning-based way in the case of a conflict-state WS-NR is purported to govern, and it is possible to drop the belief that one ought not to  $X$  in a reasoning-based way in the case of a conflict-state WS-NCR is purported to govern. After defending these claims, I will show that the source of Kolodny’s mistake lies in a faulty assumption about the starting point of reasoning in the relevant cases, which taints his formulation of the Reasoning Test.

Let us consider the following akratic scenario, purported to be governed by WS-NR. Suppose that I believe that I ought to quit smoking, but I don’t have the intention to quit smoking. Then, I reflect on my attitudes (perhaps after an epistemology lecture on subjective rationality) and come to realize the incoherence between my ought-belief and lacking the relevant intention. It seems clear that when I recognize that I am in a conflict-state like this, I can reason (*sotte voce*, as it were) as follows:

I believe that I ought to quit smoking, but I do not intend to quit smoking. This is not good because I aspire to be rational and rationality demands coherence. So, I must make my attitudes cohere with one another. Well, it seems that there are two different ways to proceed then. One is dropping my belief, and the other is forming the intention. And, I hereby decide upon doing the former.

Suppose that having reasoned thus, my ensuing decision to drop the belief in question leads to my dropping it. Call this *the akratic case*.<sup>16</sup>

It seems evident that, in this case, I have done something that is deemed impossible by the Reasoning Argument: I have dropped the belief that I ought to

---

<sup>16</sup> It might be objected that the akratic case describes an impossible scenario because belief is involuntary and therefore one cannot simply decide upon dropping a belief. There are three things I want to say in reply. Firstly, it is highly controversial that belief is involuntary, and this way of countering the akratic case makes it hostage to a controversial claim. Secondly, the objection rests on a *non-sequitur*: even if belief is involuntary and I cannot voluntarily drop a belief, I can still decide upon doing it, for instance, as a result of reasoning. Thirdly, what really matters to the plausibility of the case is not whether I voluntarily drop my belief but whether I can come to decide upon dropping my belief as a result of reasoning. And, once I decide upon dropping my belief through reasoning, that decision of mine can certainly lead to my dropping that belief, whether that way of dropping the belief counts as genuinely voluntary or not. Under such circumstances, it is plausible to say that I have dropped my belief in a reasoning-based way.



quit smoking in a reasoning-based way. And, if I can drop this belief in a reasoning-based way, as this case illustrates, then WS-NR does not collapse into NS-NR, contrary to what the Reasoning Argument attempts to show. A general lesson is that cases of akratic incoherence do not favor embellished narrow-scopism over embellished wide-scopism.

Let us now consider the following ‘converse’-akratic scenario, purported to be governed by WS-NCR. Suppose that I believe that I ought not to smoke but suppose also that I intend to smoke. Then, I reflect on my attitudes and come to realize the incoherence between my ought-not-belief and the relevant intention. It seems clear that when I recognize that I am in a conflict-state like this, I can reason as follows:

I believe that I ought not to smoke, but I intend to smoke. This is not good because I aspire to be rational and rationality demands coherence. So, I must make my attitudes cohere with one another. Well, it seems that there are two different ways to proceed then. One is dropping my belief, and the other is dropping the intention. And, I hereby decide upon doing the former.

Suppose that having reasoned thus, my ensuing decision to drop the belief in question leads to my dropping it. Call this *the converse-akratic case*.

It seems evident that, in this converse-akratic case, I have done something that is deemed impossible by the Reasoning Argument: I have dropped the belief that I ought not to smoke in a reasoning-based way. And, if I can drop this belief in a reasoning-based way, as this case illustrates, then WS-CNR does not collapse into NS-CNR, contrary to what the Reasoning Argument attempts to show. A general lesson is that cases of converse-akratic incoherence do not favor embellished narrow-scopism over embellished wide-scopism.

It is worth noting that the akratic and converse-akratic cases just described respect the two conditions Kolodny holds a process must meet in order for it to be a process of reasoning. In these cases, I move from the incoherence between my attitudes to a certain way of resolving that incoherence. Since coherence is a matter of contents of attitudes, the starting points in these cases of the mental processes are contents of my attitudes and hence Kolodny’s condition that all reasoning is content-processing is not violated. Furthermore, since the move from the incoherence between a set of attitudes to dropping one of those attitudes is clearly not upstream,<sup>17</sup> Kolodny’s other condition that upstream content-processing is not reasoning is not violated either.

---

<sup>17</sup> In the converse-akratic case, for instance, I clearly do *not* come to drop my belief that I ought not to smoke solely on the basis of recognizing that I intend to smoke. That would be progressing upstream in Kolodny’s sense of the term since it would be a move from my attitude (my

What exactly is wrong, then, with the Reasoning Argument? In both the akratic and the converse-akratic cases, I have reasoned from (the recognition of) the incoherence between my attitudes to dropping one of those conflicting attitudes, my ought-belief and my ought-not-belief, respectively. However, the Reasoning Argument assumes that the reasoning should have started from *somewhere else* if I am to be considered as having dropped the beliefs at hand in a reasoning-based way: in the akratic case, it should have started from the content of my lacking the relevant intention; and, in the converse-akratic case, it should have started from the content of my having the relevant intention. And, since, in the former case, there is no content to reason from and since, in the latter case, the move from the content of my having the relevant intention to dropping the ought-not-belief is downstream (and reasoning cannot be downstream), the Reasoning Argument concludes that one cannot come to drop those beliefs in a reasoning-based way. However, this is, I contend, unnecessarily restrictive: even if we assume, for instance, that I cannot reason, in the akratic case, from the content of my lacking the relevant intention to dropping the ought-belief or, in the converse-akratic case, from the content of my relevant intention to dropping the ought-not-belief, that by itself does not show that I cannot drop these two beliefs in a reasoning-based way. This is because I can reason from somewhere else (for instance, as both cases illustrate, from the incoherence between my belief and lacking the intention) to dropping these beliefs.

The Reasoning Test Kolodny offers is not the correct test to apply to figure out whether a proposed embellished wide scope requirement collapses into the corresponding embellished narrow scope requirement. The Reasoning Test goes, let us recall, like this:

**The Reasoning Test:** The process-requirement governing the conflict between  $X$  and  $Y$  is wide scope only if (i) it is possible to reason from the content of  $X$  to dropping  $Y$  and (ii) it is possible to reason from the content of  $Y$  to dropping  $X$ .

Even if (i) or (ii) is false, it might be possible to reason to dropping  $Y$  in a reasoning-based way and to dropping  $X$  in a reasoning-based way, in which case the process-requirement governing the conflict  $X$  and  $Y$  can still be wide scope. So, even if we grant Kolodny's observation that reasoning is required for the (thick) subjective rationality of a transition, it does not follow that the starting point of reasoning required for that transition must be the content of one or the other of

---

intention to smoke) to a reassessment of what attitude I ought to have (dropping my belief about what I ought to do). However, in the converse-akratic case, I come to drop my belief that I ought not to smoke on the basis of recognizing that that belief and my corresponding intention are incoherent. And this is not progressing upstream.

the conflicting attitudes. The Reasoning Test is a *non-sequitur* and can be plausibly rejected by embellished wide-scopism.

The adequately permissive form the reasoning test must take goes rather as follows:

**The Proper Reasoning Test:** The process-requirement governing the conflict between *X* and *Y* is wide scope only if (i) it is possible to reason (locally)<sup>18</sup> to dropping *Y* and (ii) it is possible to reason (locally) to dropping *X*.

Unlike the Reasoning Test, the Proper Reasoning Test does not place unnecessary constraints on the starting point of the required (local) reasoning process. And, as the akratic and converse-akratic cases show, WS-NR and WS-CNR pass the Proper Reasoning Test and do not thereby collapse into NS-NR and NS-CNR, respectively.

Let me conclude by replying to an objection. Consider the akratic case again, where I decide upon dropping my ought-belief, having recognized that that is one of the two things I can do to resolve the conflict-state I find myself in. According to the objection I wish to consider now, deciding to drop my ought-belief (or, simply, dropping my ought-belief) is not the subjectively rational thing to do, even if it occurs as a result of a reasoning process and I resolve the conflict by proceeding thus. If I have already judged that I ought to quit smoking, as I have done in the akratic case, I am defying this judgment by not forming the intention to quit smoking; and, defying this judgment is subjectively irrational.<sup>19</sup> So, the required reasoning process must result with forming the intention to quit smoking, if resolving the conflict-resolution in the akratic case is to count as subjectively rational – or so the objection goes.

There are two things I want to say in reply to this objection. Firstly, it is true that I defy my judgment that I ought to quit smoking by not forming the intention to quit smoking. However, wide-scopism claims that *this* is subjectively rational: in the akratic case, wide-scopism tells us, defying one's judgment is one of the possible routes that one can take in order to resolve the conflict in a subjectively rational manner. So, simply asserting that defying one's judgment is subjectively

---

<sup>18</sup> The qualification 'locally' pertains to a distinction Kolodny makes between local (subjective) requirements and global (subjective) requirements. The focus of local requirements is "specific conflicts about one's attitudes" (2005, 516), irrespective of what else is going on in her entire attitudinal system while the focus of global requirements is one's entire attitudinal system. Kolodny argues that requirements of subjective rationality "ought to be local" (2005, 516) – and hence the qualification in my formulation of the Proper Reasoning Test. It is clear that when I reason from the incoherence of a certain set of attitudes to dropping one or the other of those attitudes, my reasoning is local in Kolodny's sense of the term.

<sup>19</sup> See Kolodny (2007, 9).

irrational is question-begging as it hardly amounts to more than flatly asserting that wide-scopism is false. Secondly, (thick) subjective rationality is about maintaining coherence through reasoning that is sensitive to coherence relations; and in the akratic case, I maintain coherence by dropping my ought-belief through reasoning that is sensitive to coherence relations, which means that dropping my ought-belief in that case must be (thickly) subjectively rational. So, the objection that dropping my ought-belief is (thickly) subjectively irrational can only be properly advanced on the basis of an account that tells us what *else* is required for the (thick) subjective rationality of a transition, over and above the fact that it occurs as a result of reasoning that is sensitive to coherence relations. And, the burden clearly falls on the shoulders of the proponents of this objection.<sup>20</sup>

## References

- Bedke, M. 2009. "The iffist oughts: A guise of reasons account of end-given conditionals". *Ethics* 119(4): 672-698.
- Broome, J. 1999. "Normative requirements". *Ratio* 12(4): 398-419.
- . 2007. "Wide or narrow scope?" *Mind* 116(462): 359-370.
- Brunero, J. 2010. "The scope of rational requirements". *Philosophical Quarterly* 60(238): 28-49.
- . 2012. "Instrumental rationality, symmetry, and scope". *Philosophical Studies* 157: 125-140.
- Darwall, S. 1983. *Impartial reason*. Cornell UP.
- Fogal, D. & A. Worsnip. 2021. "Which reasons? Which rationality?" *Ergo* 8(11): 306-343.
- Goldman, A. 1979. "What is justified belief?" In *Justification and Knowledge*, edited by G. S. Pappas, 1-23. D. Reidel Publishing Company, Dordrecht: Holland.
- Greenspan, P. 1975. "Conditional oughts and hypothetical imperatives". *Journal of Philosophy* 72: 59-276.
- Hill, T. 1973. "The hypothetical imperative". *Philosophical Review* 82(4): 429-450.
- Huemer, M. 2011. "The puzzle of metacoherence". *Philosophy and Phenomenological Research* 82(1): 1-21.
- Kolodny, N. 2005. "Why be rational?" *Mind* 114(455): 509-563.
- . 2007. "State or process requirements?" *Mind* 116(462): 1-15.
- Lord, E. 2014. "The real symmetry problem(s) for wide-scope accounts of rationality". *Philosophical Studies* 170: 443-464.

---

<sup>20</sup> The author states that there is no conflict of interest.

- Schroeder, M. 2004. "The scope of instrumental reason". *Philosophical Perspectives* 18: 337-364.
- . 2009. "Means-end coherence, stringency, and subjective reasons". *Philosophical Studies*, 143(2): 337-364.
- Silva P. and L. Oliveira (eds). 2022. *Propositional and doxastic justification: new essays on their nature and significance*. Routledge: New York.
- Wallace, R. 2001. "Normativity, commitment and instrumental reason". *Philosophers' Imprint* 1, 1-26.
- Way, J. 2011. "The symmetry of rational requirements". *Philosophical Studies* 155: 227-239.
- Worsnip, A. 2018. "The conflict of evidence and coherence". *Philosophy and Phenomenological Research* 96(1): 3-44.



# CONTEXTUAL SHIFTS AND GRADABLE KNOWLEDGE

Andreas STEPHENS

**ABSTRACT:** Epistemological contextualism states that propositions about knowledge, expressed in sentences like “S knows that P,” are context-sensitive. Schaffer (2005) examines whether one of Lewis’ (1996), Cohen’s (1988) and DeRose’s (1995) influential contextualist accounts is preferable to the others. According to Schaffer, Lewis’ theory of relevant alternatives succeeds as a linguistic basis for contextualism and as an explanation of what the parameter that shifts with context is, while Cohen’s theory of thresholds and DeRose’s theory of standards fail. This paper argues that Schaffer’s analysis is unsatisfactory since it fails to show that thresholds and standards cannot cope with skepticism, as it is ultimately the conversation participants who control how the conversation plays out. Moreover, Schaffer fails to show that gradability is of no importance in inquiries.

**KEYWORDS:** knowledge, contextualism, Schaffer, thresholds, standards, alternatives, gradability, gradualism

## 1. Introduction

Epistemological contextualism emphasizes the context-sensitivity of epistemological concepts such as ‘knowledge’ (see, e.g., Lewis 1979; DeRose 2009; Rysiew 2016). In this paper, I will focus on Schaffer’s (2005) analysis of whether one of Cohen’s (1988), DeRose’s (1995), and Lewis’ (1996) influential contextualist accounts can explain what parameter it is that shifts between contexts in knowledge-ascriptions. In short, the three theories suggest different parameters: Cohen suggests that it is whether the threshold value for full justification is reached; DeRose suggests that it is whether the strength of the epistemic position given a particular standard is sufficient, and; Lewis suggests that it is whether relevant alternatives are possible to eliminate.

Schaffer concludes that Lewis’ theory of alternatives provides a workable explanation of what the parameter that shifts between contexts is, whereas Cohen’s and DeRose’s theories fail. I will problematize Schaffer’s conclusion and raise two issues to shed light on how Cohen’s and DeRose’s theories seemingly can, after all, describe certain aspects of ‘knowledge’ that Lewis’ theory does not account for. I will argue that Schaffer fails to convincingly show that Cohen’s and DeRose’s

theories cannot cope with skepticism. Moreover, Schaffer fails to show that thresholds and standards do not provide relevant input for inquiries.

It should be pointed out that other theories such as, for example, subject sensitive invariantism and relativism will not be discussed. Furthermore, I will not address Schaffer's contrastivist position or his wider discussion of contextualism, which he details in several other papers (see, e.g., Schaffer 2001, 2004, 2007, 2008; Schaffer and Knobe 2012; Schaffer and Szabó 2014). Instead, I will engage with his lucidly argued (2005) paper, although Schaffer also addresses this specific topic in other texts (see, e.g., Schaffer 2006, section 4, 2007, sections 2 and 5, 2015, section 30.1.3; Schaffer and Szabó 2014, section 2.2).

## 2. Schaffer's analysis of contextualism

To investigate what the parameter that shifts with context is, Schaffer (2005, 116-118) lists several requirements that an answer should be able to meet. These are given in the form of four desiderata or criteria.

I. *Linguistically plausible*. The parameter should be a naturally occurring linguistic parameter and work together with the concept of knowledge. That is, the parameter cannot be invented just to serve this purpose but should exist naturally in language and be applicable to knowledge and similar concepts.

II. *Predictively adequate*. The parameter should follow our intuitions for what is reasonable for knowledge ascriptions. The parameter should therefore not be linked to anything else that may cause the shifts. Thus, for example, which day it is, must not affect the parameter (Schaffer 2005, 116).

III. *Skeptically resolving*. The parameter should enable problems with various forms of skepticism to be solved in accordance with contextualism. The parameter should thus enable a contextualist explanation that 'renders most ordinary knowledge ascriptions true in ordinary contexts, some (those associated with the specific doubt in play) false in moderately skeptical contexts, and most (or perhaps all) false in radically skeptical contexts' (Schaffer 2005, 117).

IV. *Illuminate inquiry*. The parameter should shed light on the practical role of knowledge ascription for our investigations. For Schaffer, this means that knowledge ascriptions involve that a subject S can answer questions, which Schaffer illustrates through three possible inquiries with three different questions (Schaffer 2005, 117):

- (i): "Is there a goldfinch in the garden, or a blue jay?"
- (ii): "Is there a goldfinch in the garden, or a canary?"
- (iii): "Is there a goldfinch in the garden, or at the neighbor's?"



According to Schaffer, the knowledge ascription “I know that there is a goldfinch in the garden” has different roles in the three different contexts. In the first context, the ascription of knowledge means that the subject can distinguish a goldfinch from a blue jay. In the second context, the ascription of knowledge means that the subject can distinguish a goldfinch from a canary, which is a more difficult task than the task in the first context. In the third context, the ascription of knowledge means that the subject can distinguish his garden from the neighbor’s, which is a completely different kind of task than those in the first and second contexts (Schaffer 2005, 117-118). Schaffer states that the parameter that is being shifted should be able to explain how this is done.

As I view them as reasonable starting points, I will not focus on criticizing Schaffer’s choice of criteria. Now, as aforementioned, Schaffer addresses three possible parameters for contextual shifts: a shift of the *threshold* for full justification; a shift in the *standard* for an epistemic position, and; a shift in the epistemic *alternatives* (Schaffer 2005, 115-116), which we will turn to below.

## 2.1 Thresholds

Schaffer bases his understanding of how thresholds might be the parameter that shifts between contexts on Cohen (1988). The starting point is described to be the threshold value (T) for full justification. In Schaffer’s interpretation, this is presented as S being ascribed an absolute degree of justification (D) for P. What shifts between different contexts is whether D is sufficient to be considered fully justified. D’s sufficiency is controlled by whether D reaches T (Schaffer 2005, 118). The degree of justification can then be compared to an interval between 0 and 1, where movements take place up or down the interval. The context can select a T in different places on the interval. If S’s absolute justification D in a context reaches the threshold value T ( $D \geq T$ ), S is fully justified and thus knows. If, on the other hand, there is a change in contexts that moves the threshold value T so that S’s justification D does not reach T ( $D < T$ ), S does not know. Schaffer’s interpretation of Cohen strikes me as reasonable.

Schaffer then examines whether thresholds can meet the criteria (I–IV) he lists, and concludes that thresholds fail on all accounts. Regarding his first criterion, Schaffer believes that the parameter does not work for the concept of knowledge. Rather, the parameter works for gradable adjectives such as ‘tall’ and ‘justified.’ But ‘knowledge’ is neither an adjective nor gradable according to Schaffer. He further argues that knowledge can be considered to imply justification, but claims that this fact in itself does not make knowledge vague or gradable (Schaffer 2005, 119).

In relation to the second criterion, Schaffer believes that thresholds predict incorrect shifts. As the requirement for justification shifts, we would, according to Schaffer, shift our view of whether knowledge exists or not. This could take place by using comparison classes or by drawing a line (Schaffer 2005, 119). But, in Schaffer's view, such comparison classes do not seem to affect any shifts in the concept of knowledge and what we know, it is instead specific doubts that make us shift our knowledge ascriptions.

Schaffer believes that thresholds do not meet the third criterion concerning skepticism either. Partly because it is not clear why skeptical doubts would create any shifts at all for the threshold value. So, it is unclear why doubts about 'brain-in-a-vat scenarios' should affect the threshold value – make it really high. The skeptic's mentioning of something else than what was originally in focus does not explain why this would affect the threshold value for what we first focused on. Schaffer's second reason for why thresholds do not meet his third criterion is that doubt would shift the threshold value in the wrong way for the skeptic who doubts something specific. When the threshold value shifts, it does so in general, not just regarding something specific, and all truth values are thus affected in that context. But in some skeptical scenarios this does not happen. Statements will be ascribed to the absolute degree of justification,  $D1$  and  $D2$ , and it seems plausible, in Schaffer's view, that  $D1 = D2$ . But in that case the threshold value is raised in both cases if it is raised in one. If, on the other hand,  $D1 \neq D2$ , the lower threshold value cannot be raised above the higher without affecting the other. This means that if you raise the justification requirement somewhere, it affects the justification requirement everywhere. The interconnection is thus problematic for the threshold-parameter (Schaffer 2005, 120).

Finally, Schaffer argues that thresholds fail to account for the fourth criterion, as there need not be a connection between investigations and the threshold value for justification. Schaffer questions that there needs to be a link between the justification threshold for the statement "I know there is a goldfinch in the garden" and investigations into whether this is in relation to a blue jay; a canary, or; the neighbor's garden. The threshold value would then be raised everywhere, similarly as concerning the third criterion. Although there is a difference between justification regarding different bird species and whether the bird is in my or the neighbor's garden, the threshold theory according to Schaffer has the consequence that raising the threshold for justification concerning one question affects the threshold value for all questions.

Schaffer argues that the interconnection makes the theory plausible. But what shifts between contexts must be more sensitive to distinct doubts, which

thresholds fail to be (Schaffer 2005, 119-121). Although I will question his conclusion below, Schaffer does present a lucid and strong case against thresholds.

## 2.2 Standards

Schaffer describes how standards, based on DeRose (1995), shows how what shifts between contexts is how strong an epistemic position needs to be for knowledge. So, S can be ascribed a specific absolute strength for his epistemic position R regarding his belief P. R is likened to the radius of a sphere of possible worlds where the same metric M is used. M is determined by the context, where S can follow the truth and falsehood of P (track the truth). If R reaches far enough, described as a standard radius L, given M, S knows. What differs between contexts is both the kind of sphere of possible worlds that is relevant and whether R reaches far enough for knowledge (Schaffer 2005, 121). The context thus determines M and determines whether  $R \geq L$  (there is knowledge given M). In cases where  $R < L$  there is no knowledge (Schaffer 2005, 121; DeRose 2009, 14). M can expand or shrink depending on the context.

Schaffer describes an example where S can follow the truth given the starting point x and the sphere w1, while this is not the case given the sphere w2. x is thus a central starting point and w1 and w2 are spheres around x. If context 1 gives that M is  $\langle x, w1, w2 \rangle$  and L is 1, then S's belief that P is knowledge in context 1, where S can follow the truth as far as the context requires. R covers x and w1, and L indicates that only one step from x is necessary for knowledge. In context 2 where M is the same as in context 1 but L is 2, R no longer reaches far enough to give S knowledge that P, since S cannot follow the truth given w2. If M in context 3 instead is  $\langle x, w2, w1 \rangle$  and L is 1, then S is not considered to know that P since S cannot follow the truth given w2. In context 3, S can thus only reach one step from x and then reach w2, but S can, as previously mentioned, only follow the truth given w1. The parameter that shifts between contexts is then what kind of sphere of possible worlds is relevant and the standard for how far S must follow – track the truth – for knowledge (Schaffer 2005, 121).

Schaffer mentions how standards are reminiscent of thresholds, a point I agree with. As far as S needs to follow the truth (R), on DeRose's take, can correspond to Cohen's degree of justification (D) (Schaffer 2005, 121-123). However, the theories differ in that D may have an upper limit at 1 while R does not have an upper limit. Standards also include an extra parameter in the form of the concept of sphere of possible worlds (Schaffer 2005, 122).

Schaffer argues that standards – just as thresholds – fail concerning all four criteria. Since standards have not specified any general parameter linked to

knowledge that is usually found in our language, they do not meet the first criterion. There is thus nothing that can function as a precursor to the parameter in our language. The parameter instead, in his view, seems to be taken out of thin air, which the first criterion does not allow (Schaffer 2005, 123).

Regarding the second criterion, Schaffer argues that the wrong shifts are predicted by standards, as shifts regarding what we know and what we do not know should occur when  $M$  is changed and when  $L$  is raised or lowered. In short, this can be described as when the context shifts the strength required for knowledge of an epistemic position is increased or decreased. Schaffer believes, however, that the connection between which the possible worlds are in the context does not have to affect what we count as knowledge. As mentioned earlier regarding thresholds, what makes us shift between situations where we consider ourselves to know and not know is, according to Schaffer, the introduction of particular doubts.

Schaffer believes that standards cannot provide a reasonable contextualist explanation for skeptical problems as it is unclear why skeptical problems would cause any shift in standard. If standards are changed, they would, through 'the standard theory,' affect the truth of propositions, in skeptical problems that focus on a specific thing in the wrong way (Schaffer 2005, 124). The spheres, governed by context, would take into account too much to function. As with thresholds, Schaffer sees the connection as problematic.

Finally, regarding the fourth criterion, Schaffer believes that standards fail, as questions will be affected and controlled by the scenario that is furthest away, in relation to possible worlds. Regardless of the conditions, the most remote scenario will drive up the standard for other scenarios as well. In Schaffer's view, the spheres DeRose uses are too limited to be able to describe what we need. According to Schaffer, standards are too interconnected and do not allow the measure of independent possibilities required for a plausible parameter.

Again, Schaffer presents a strong case against the parameter in question.

### 2.3 Alternatives

The third parameter addressed by Schaffer states that what shifts in knowledge ascriptions in different contexts is a quantity consisting of which epistemic alternatives  $S$  must take into account and which alternatives  $S$  can eliminate. Schaffer links this account to Lewis (1996). Schaffer describes how  $S$ 's belief that  $P$  is ascribed to an absolute elimination force  $E$  for  $P$ , where  $E$  is the number of possibilities  $S$  can eliminate. Which, and how many, relevant alternatives are sufficient, shifts between contexts and the amount of them is ascribed value  $Q$ .

Knowledge arises when E covers Q and S thus has eliminated a sufficient number of alternatives (Schaffer 2005, 125). To illustrate this, according to Schaffer, E can be seen as an arbitrary region and the relevant alternatives Q as another arbitrary region. What shifts is whether E covers Q. To summarize the theory's view of shifts between contexts: The parameter that shifts between contexts is which, and how many, options S must eliminate (Schaffer 2005, 125).

Schaffer gives an example where the regions are likened to a chessboard where P is a certain square. Q stands for the other squares on the board that are relevant options. E stands for the squares S can remove and covers the black squares. If context 1 indicates that Q (the relevant alternatives) only apply to black squares, S has knowledge that P in context 1 since E covers Q. If context 2 indicates that Q applies to the white squares, and context 3 that Q applies to both black and white squares, S does not know that P in any of these contexts, as E in those cases does not cover Q (Schaffer 2005, 125-126).

The reason why the alternatives should be seen as arbitrary regions is that Schaffer wants to point out that nothing connects the alternatives, which should instead be seen as completely independent. Relevant alternatives thus do not need to have a direct connection between them. This also means that no grading between alternatives is possible. An example of two independent alternatives can be given through an ascription of knowledge about knowing what it means to build a computer. "S knows what it means to build a computer" can mean that S knows that certain components are needed to assemble a computer. An independent alternative to the ascription of knowledge can instead mean a completely different kind of thing, for example, that S knows that it means being able to draw all the parts needed in a CAD program. Schaffer's point is that there is no limit to what the alternatives can mean and how different alternatives may be. Radical skepticism's alternatives of, for example, brains-in-vats are extreme cases of alternatives that differ from the alternatives we usually consider. Nothing thus links the alternatives directly to each other. This aspect was elucidated by Schaffer in his fourth criterion.

Schaffer investigates whether the theory focused on alternatives is able to meet the necessary criteria he raised. His first criterion requiring that the parameter is naturally occurring in language can be seen to be fulfilled as the concept of knowledge can be compared to the concepts of 'can,' 'must' and 'regret.' In his view, statements such as "I can run a mile in ten minutes" and "wood must burn" are context-sensitive (Schaffer 2005, 126). Whether these statements are true or not thus depends on what alternatives exist and are relevant. In a context where the relevant alternatives include that I am an incredibly good runner, and that

wood is always burning, the statements can be true. If, on the other hand, I am a bad runner and other laws of nature than those we know are relevant alternatives, the statements can be false. According to Schaffer, this means that knowledge can be linked to a naturally occurring parameter for alternatives.

What shifts between cases where we “have knowledge” and “do not have knowledge” thus seems to be due to special doubts. If the introduction and removal of these doubts are understood in terms of relevance for specific alternatives, Schaffer’s second criterion seems to be fulfilled (Schaffer 2005, 127).

In relation to his third criterion, an explanation is needed as to why skepticism can create particular doubts and how radical skepticism can make us doubt in general. A theory focused on alternatives seems to be able to explain both of these forms of doubt through how the amount of alternatives is expanded. That I am a brain-in-a-vat, which the radical skeptic often takes as an example, is a relevant alternative to basically all the facts I can point out about the outside world. A milder form of skepticism where the alternative that my car has been stolen is mentioned might create doubt concerning that question but not concerning the unrelated question of when the theater opens. This option also does not affect any of the other everyday facts I take for granted. This is an important point to highlight in Schaffer’s overall argumentation as this is a main reason for why alternatives differ from thresholds and standards. The ability of alternatives to be independent is the reason why Schaffer considers them a good explanation for what shifts between contexts.

In Schaffer’s fourth criterion a parameter was needed that could illuminate our investigations. Alternatives can fulfill this role as questions are always, or can be formulated as, multiple-choice questions. The sets of alternatives can be seen as discrete/independent and can thus explain the questions mentioned above. The relevant alternatives will be: {goldfinch in the garden, blue jay in the garden}; {goldfinch in the garden, canary in the garden}; {goldfinch in the garden, goldfinch in the neighbor’s garden}. The alternatives thus differ in what they presuppose (Schaffer 2005, 128).

As mentioned initially in this section, Schaffer believes that alternatives do serve as a reasonable description of which parameter it is that shifts between contexts and thus gives contextualism a plausible starting point as a theory of knowledge. According to Schaffer, the independent property of alternatives is the main reason for his assessment.

## 2.4 Schaffer's conclusion

Schaffer claims to have shown how alternatives provide a working description of what the parameter that shifts between contexts is. In contrast, both thresholds and standards fail to describe what shifts between different contexts. The crucial point that causes Schaffer's conclusion is that alternatives enable independent options, while thresholds and standards are interconnected. Thresholds in the form of points on a range of justification and standards in contextual spheres and a radius of strength of epistemic position coupled to the outer edge of the spheres. Alternatives thus enable options without internal connection. Thresholds and standards enable gradability.

But is the gradability of thresholds and standards really completely irrelevant to depict what shifts between contexts in our knowledge ascriptions?

## 3. Two issues with Schaffer's analysis

Schaffer concludes that alternatives can explain which parameter it is that shifts between contexts by fulfilling the four criteria he presents – thresholds and standards instead fail with all of them. I will question Schaffer's analysis of the third and fourth criteria (concerning skepticism and the practical role in investigations) arguing that there are aspects of these criteria that contradict Schaffer's conclusion. Thresholds and standards seem to cope better with skepticism than Schaffer's analysis claims, and they seem to depict something in our practical investigations which alternatives miss. So, thresholds and standards seemingly can – contrary to Schaffer's opinion – offer relevant input concerning the parameter that shifts between contexts, and thus provide insight into the concept of knowledge. The third and fourth criteria are in focus since Schaffer places the greatest emphasis on them in his analysis, and because the first two (concerning linguistic plausibility and appropriate predictability) are affected by one's view of the latter. For greater clarity, some repetitions of previous paragraphs will be used below.

### 3.1 Thresholds and standards can cope with skepticism

First a quick recap. The parameter-shifts between contexts must be able to tackle skepticism in a manner that is in accordance with contextualism. The parameter must thus enable a contextualist explanation of how: ordinary knowledge ascriptions can be true in everyday contexts; certain knowledge ascriptions may be false in skeptical contexts where specific doubts have been raised; all knowledge ascriptions can be false in radical skeptical contexts (Schaffer 2005, 116-117).

Schaffer believes that neither thresholds nor standards meet the requirement to work as contextualist solutions to skeptical problems. He argues that it is unclear why skeptical doubts would create any shifts at all in terms of thresholds or standards. Why should doubt about whether I am a brain-in-a-vat affect the threshold or the standard? He also claims that raised doubts would affect the threshold and standard incorrectly for moderate skeptics. I will question the plausibility of Schaffer's view.

Schaffer's treatment of thresholds and standards is very similar and two main points are used in both cases:

(a<sup>1</sup>): It is unclear why skeptical arguments should affect us at all (our thresholds or standards);

(a<sup>2</sup>): Doubts affect the threshold and standard incorrectly for moderate skeptics.

The first point, (a<sup>1</sup>), can be questioned by pointing out how it is the conversation participants who decide whether, when, and how a skeptical argument applies or not. The answer to (a<sup>1</sup>) thus depends on whether the conversation participants accept the skeptic's argument or not. This provides support for thresholds and standards as the appearance of ambiguity that Schaffer's objection stipulates and relies on is entirely up to the conversation participants to handle. This thus applies not only to both thresholds and standards but also to alternatives. What Schaffer misses is that how we use thresholds for full justification or standards – as well as alternatives – is governed by the conversation participants and it is up to *them* how far the threshold or standard should extend. Assuming that this is "unclear" is thus unreasonable.

This view is in line with how Lewis writes that what one may assume and ignore is determined by speakers and listeners in a context (Lewis 1996, 378-379). Another passage from Lewis (1979) elaborates his position:

At any stage in a well-run conversation, a certain amount is presupposed. The parties to the conversation take it for granted; or at least they purport to, whether sincerely or just 'for the sake of the argument'. Presuppositions can be created or destroyed in the course of a conversation. (Lewis 1979, 339)

Cohen (1988) describes how the skeptic *forces our attention* and it is when we fail to realize this that we experience that we have problems with the skeptic's argument. Thus, even if the skeptic's argument is based on compelling rhetoric, we do not *have to* agree with, or accept, the skeptic's argument:

Skeptical arguments exploit the fact that certain considerations can lead to a shift in the standards of relevance. Failure to recognize the shift can lead us into paradox. [...] The apparent closure failures are illusions that result from



inattention to contextual shifts. (Cohen 1988, 110-111)

Finally, DeRose leaves open the question of whether the skeptic actually manages to shift the context:

For the fact that the skeptic can invoke very high standards that we do not live up to has no tendency to show that we do not satisfy the more relaxed standards that are in place in more ordinary conversations and debates. (DeRose 1995, 5)

Thus, according to DeRose, it is rather a question of choice, as pointed out above, the conversation participants make about how the conversation should proceed and what they want to communicate. DeRose's pragmatic argument for why we should prefer contextualism over skepticism is that we seem to want knowledge. The important thing to highlight is that we can choose:

Indeed, since the bold skeptical solution and our new contextualist solution under consideration closely parallel each other, there's not much difference in how they solve the puzzle. That the bold skeptical resolution involves us in systematic falsehood is one of the few differences to be found here, and it's a weighty consideration against that resolution. (DeRose 1995, 48-49)

The lines of inquiry that have been addressed illustrate how thresholds and standards do not need to have a problem with skepticism if one starts from Lewis', Cohen's and DeRose's views of what happens in a conversation rather than Schaffer's. Thus, it does not have to be unclear whether, and if so when or how, skeptical arguments should affect us – it is up to the conversation participants. (Lewis' (1979, 340-341) 'rule of accommodation for permissibility' highlights some relevant aspects to how there tends to be hierarchical differences that govern conversations.) In relation to (a<sup>1</sup>), and Schaffer's third criterion, thresholds, standards, and alternatives can thus be considered equally plausible in relation to skepticism.

Concerning (a<sup>2</sup>), Schaffer's initial formulation is problematic. When he exemplifies the skeptical scenarios that are introduced for a subject, he formulates the contexts as follows:

(i) 'I know that my car is parked on Elm', and (ii) 'I know that the movie starts at nine.' In a moderately skeptical context in which unresolved doubts have been raised as to whether my car has been stolen and relocated, (i) should count as false, though (ii) should still count as true (no doubts have yet been raised about *that*). Whereas in a radically skeptical context in which unresolved doubts have been raised as to whether one is dreaming, or a brain-in-a-vat, etc., (i) and (ii) should both count as false. (Schaffer 2005, 117, italics in original)

This, in my view, tips the scales in favor of the alternatives-account and is not a neutral formulation. On the one hand, if one insists on an interpretation, as

Schaffer does, where threshold- and standard-shifts '*globally infect* other truth-values' (Schaffer 2005, 119, italics in original) it might be correct that the theory of alternatives offers a better explanation – since the presented formulation specifically involves discrete/independent relevant alternatives. Regarding (a<sup>2</sup>), thresholds and standards might then indeed not be optimal (and possibly incorrect) for explaining moderate skeptical scenarios. If threshold- and standard-theories are disallowed to invoke any sensitivity to different “epistemic positions” they could possibly fail to make sense of this specific matter. However, by reformulating the case set-up slightly, highlighting gradable qualities of knowledge, other intuitions might arise that instead pose problems for alternatives-theories. That is, if the shift in context, and the introduced doubts, is formulated to involve *how much* or *how well* I know about the relevant topics, thresholds and standards might fare better than alternatives at explaining the shifts. This point will be elaborated on below concerning the second issue.

### 3.2 Alternatives misses an aspect of knowledge ascriptions that thresholds and standards can explain

Schaffer's fourth criterion pointed out that the parameter that shifts between contexts should shed light on the practical role of knowledge ascriptions for our investigations. Knowledge ascriptions to the subject S thus mean, according to Schaffer, that S can answer questions. These questions are similar to the parameter for knowledge ascription linked to different contexts. The parameter must be able to illuminate and evaluate these issues (Schaffer 2005, 117-118). Schaffer uses three different examples of contexts and questions in them:

- (i): “Is there a goldfinch in the garden, or a blue jay?”
- (ii): “Is there a goldfinch in the garden, or a canary?”
- (iii): “Is there a goldfinch in the garden, or at the neighbor's?”

In context (i), the ascription of knowledge means that the speaker can distinguish a goldfinch from a blue jay. In context (ii), the ascription of knowledge means that the speaker can distinguish a goldfinch from a canary, which is more difficult (Schaffer 2005, 117). In context (iii), the ascription of knowledge means that the speaker can distinguish his garden from the neighbor's, which is a completely different kind of task (Schaffer 2005, 117-118). The parameter that shifts should be able to explain how this is done.

In his answer, Schaffer wants to show that the fourth criterion is fulfilled through the parameter of alternatives, as questions are always, or can be, formulated as multiple-choice questions. The sets of alternatives can thus be seen

as discrete/independent and can hence explain the questions in (i)–(iii), or so Schaffer argues (Schaffer 2005, 128). Thresholds and alternatives instead fail since:

(b): Thresholds and standards are too interconnected (in how they treat increases and decreases of the threshold value and the standard, respectively).

But alternatives only seem to fully capture what happens in the shift between context (ii) and (iii), where another kind of thing constitutes the relevant alternative. In the shift between context (i) and (ii), alternatives instead miss an aspect of what we are expected to know, or so I claim. Schaffer himself described how (ii) means that something *more difficult* needs to be done. If the alternatives are independent, which Schaffer uses as a reason to consider that alternatives are the preferred parameter, it also means that the aspect that someone who can distinguish between a goldfinch and a canary can do something *more/better* than someone who can only distinguish between a goldfinch and a blue jay is ignored. Here, the interconnection of thresholds and standards, i.e., gradability and possibilities of increases and decreases in strength, seems to be just what is sought after. The *interconnection* pointed out in (b), which Schaffer raised as an objection to thresholds and standards, thus seems to constitute exactly what is needed to illuminate the aspect of the concept of knowledge that we are looking for.

The interconnection can be seen in both Cohen's thresholds and DeRose's standards. In these cases, it is gradability that is in focus, which shows that the connection that Lewis and Schaffer want to remove from the concept of knowledge – what Cohen and DeRose take into account – is exactly what is sought after. DeRose explains how we can shift our epistemic positions and how it can be required more or less/something better or worse of us:

[...] One could gather further evidence, strengthen one's epistemic position with respect to both not-H and O, and make even one's belief that not-H sensitive.  
(DeRose 1995, 33)

Schaffer and Szabó (2014) discuss this matter in connection to comparing knowledge with gradable adjectives and verbs. They conclude that there is a precedent for using 'knows' in a gradable manner, although they stress that it is in an idiomatic and *ad hoc* way (Schaffer and Szabó 2014, 503). They also argue that 'perhaps there is literal grading, but of something other than the knowledge state.' (Schaffer and Szabó 2014, 504, fn. 14). However, other interpretations seem possible. For example, Dutant (2007) suggests that knowledge can involve degree modifiers and Lai (2019, 6, italics in original) argues that knowledge should be interpreted '[...] as a *spectrum concept* analogous to 'red', 'bright', and 'cold'.' (see

also, e.g., Hetherington 2001). Such gradualism thus highlights that propositions can be known more or less/better or worse.

The interconnection and the gradability of the concept of knowledge, as well as justification, thus seem to be needed to depict all aspects of Schaffer's fourth criterion. Thresholds, as well as standards, thus seem to be able to add something to the investigation of the concept of knowledge that alternatives miss. Schaffer's second point (b) thus turns out to highlight something that proves to be a strength with thresholds and standards – contrary to what Schaffer claims.

#### 4. Concluding remarks

Schaffer fails to show that thresholds and standards have a problem with skepticism, as it is ultimately the conversation participants who control how the conversation plays out. It is the conversation participants who decide whether, and if so how, the interconnection in the theories affects the conversations and whether this should be seen as a problem or not. Moreover, Schaffer's analysis fails to show that gradability is of no importance concerning his fourth criterion – which thresholds and standards account for better than alternatives.

Some type of gradualism, where knowledge is seen as a spectrum concept and/or propositions are seen as being possible to know more or less/better or worse, thus seemingly remains a plausible option. Now, since (human) communication is a complex natural phenomenon, and all models must involve *some* abstraction and idealization, it also seems reasonable – indeed inevitable – that any particular parameter will only give a partial picture.<sup>1</sup>

#### References

- Cohen, S. 1988. "How to be a fallibilist." *Philosophical Perspectives*. 2 (Epistemology): 91-123.
- DeRose, K. 1995. "Solving the skeptical problem." *The Philosophical Review* 104(1): 1-52.
- . 2009. *The case for contextualism: Knowledge, skepticism, and context, Vol. 1*. Oxford: Oxford University Press.
- Dutant, J. 2007. "The limited gradability of 'knows'". *unpublished manuscript*.
- Hetherington, S. 2001. *Good knowledge, bad knowledge: On two dogmas of epistemology*. Oxford: Oxford University Press.

---

<sup>1</sup> I am grateful for Stefan Schubert's knowledgeable input concerning an earlier version of this text. Thanks to Asa Kasher, Martin L. Jönsson and my anonymous reviewers for comments.

- Lai, C. 2019. "The concept of gradable knowledge." The University of Edinburgh. *Doctoral thesis*.
- Lewis, D. 1979. "Scorekeeping in a language game." *Journal of Philosophical Logic* 8: 339-359.
- . 1996. "Elusive knowledge." *Australasian Journal of Philosophy* 74(4): 549-567.
- Rysiew, P. 2016. "Epistemic contextualism." In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by E. N. Zalta. URL = <<https://plato.stanford.edu/archives/win2016/entries/contextualism-epistemology/>>.
- Schaffer, J. 2001. "Knowledge, relevant alternatives and missed clues." *Analysis* 61(271): 202-208.
- . 2004. "From contextualism to contrastivism." *Philosophical Studies* 119(1/2): 73-103.
- . 2005. "What shifts? Thresholds, standards, or alternatives?." In *Contextualism in philosophy: Knowledge, meaning, and truth*, edited by G. Preyer and G. Peter, 115-130. Oxford: Oxford University Press.
- . 2006. "The irrelevance of the subject: Against subject-sensitive invariantism." *Philosophical Studies* 127(1): 87-107.
- . 2007. "Knowing the answer." *Philosophy and Phenomenological Research*, 75(2): 383-403.
- . 2008. "The contrast-sensitivity of knowledge ascriptions." *Social Epistemology* 22(3): 235-245.
- . 2015. "Lewis on knowledge ascriptions." In *A companion to David Lewis*, edited by B. Loewer and J. Schaffer, 473-490. Oxford: John Wiley & Sons, Inc.
- Schaffer, J., and Knobe, J. 2012. "Contrastive knowledge surveyed." *Noûs* 46(4): 675-708.
- Schaffer, J., and Szabó, Z. G. 2014. "Epistemic comparativism: A contextualist semantics for knowledge ascriptions." *Philosophical Studies* 168(2): 491-543.



## NOTES ON THE CONTRIBUTORS

**Frederik J. Andersen** is a PhD student at Arché, Philosophical Research Centre for Logic, Language, Metaphysics, and Epistemology, University of St Andrews. His main research interests are in epistemology and logic. In September 2023 he will submit his doctoral project on Logical Disagreement. In 2024 he is expected to publish a co-authored entry on the same topic with Professor Anandi Hattiangadi for the *Oxford Handbook of Philosophy of Logic*, Oxford University Press. Contact: frederikjandersen@gmail.com.

**Arnold Cusmariu** is an analytic philosopher with a PhD from Brown University. He has published articles on metaphysics, epistemology, logic, the philosophy of language and aesthetics, available online at <https://www.researchgate.net/profile/Arnold-Cusmariu>. His book *Logic for Kids* is available online at Amazon, Barnes & Noble, and Books-a-Million. Contact: bravo323@gmail.com.

**Danilo Fraga Dantas** is a Brazilian philosopher who has earned his Ph.D. in Philosophy from the University of California, Davis. He currently serves as a professor of Logic and Epistemology at the Federal University of Paraíba, in Brazil. Dantas employs computer simulations to explore how formal models of rationality can account for the cognitive limitations of finite reasoners. His work has been featured in reputable academic journals, including “How to (Blind)Spot the Truth: An Investigation on Actual Epistemic Value” in *Erkenntnis* and “Epistemic Sanity or Why You Shouldn’t be Opinionated or Skeptical” in *Episteme*. For inquiries and collaboration opportunities, you can reach Danilo Dantas at [dfdantas@ucdavis.edu](mailto:dfdantas@ucdavis.edu).

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

**Andreas Stephens** is currently studying for a PhD in theoretical philosophy at Lund University. His research is focused on the intersection of naturalistic epistemology and cognitive science. Contact: [andreas.stephens@fil.lu.se](mailto:andreas.stephens@fil.lu.se).





# 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](mailto:logosandepisteme@yahoo.com). Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

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

## 2. Publication Ethics

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

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

## 3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

#### **4. Manuscript Format**

Manuscripts should be formatted in Rich Text Format file (\*.rtf) or Microsoft Word document (\*.docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

#### **5. Official Languages**

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

#### **6. Abstract**

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

#### **7. Author's CV**

A short CV including the author's affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

#### **8. Review Process**

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in

their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

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

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

## **9. Acceptance of the Papers**

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

## **10. Responsibilities**

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

## **11. Checking Proofs**


Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

## **12. Reviews**

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice și Sociale „Gh. Zane” Academia Română, Filiala Iași, Str. Teodor Codrescu, Nr. 2, 700481, Iași, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

### **13. Copyright & Publishing Rights**

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License () . Authors have the right to use, reuse and build upon their papers 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.

No manuscripts will be returned to their authors. The journal does not pay royalties.

### **14. Electronic Archives**

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

## ***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 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](mailto:logosandepisteme@yahoo.com).