Synthese (2022) 200:425
https://doi.org/10.1007/s11229-022-03905-6

ORIGINAL RESEARCH

On the pure logic of justified belief

Daniela Schuster[1] · **Leon Horsten**[1]

Received: 5 August 2021 / Accepted: 16 August 2022 / Published online: 12 October 2022

Abstract
Justified belief is a core concept in epistemology and there has been an increasing interest in its logic over the last years. While many logical investigations consider justified belief as an operator, in this paper, we propose a logic for justified belief in which the relevant notion is treated as a predicate instead. Although this gives rise to the possibility of liar-like paradoxes, a predicate treatment allows for a rich and highly expressive framework, which lives up to the universal ambitions of investigating epistemological concepts. We start with a base theory for justified belief, and then systematically present putative additional axioms for justified belief. We provide an overview of (in)consistency results when the additional principles are added to the base theory, and discuss their philosophical plausibility.

Keywords Justified belief · Paradox · Logic of justification · Self-reference

1 Introduction

Justified belief is a fundamental epistemological concept. Even though there is deep disagreement among epistemologists about the content of the concept of justified belief, most philosophers believe that there is a core concept of justified belief that this disagreement is about. In this paper, we are concerned with the logical principles that govern this core concept.

Over the past two decades, the logic of justification has been more intensively investigated than before. It has been investigated in different frameworks and in different ways. Often a multi-modal framework is adopted, in which the notion of justified belief is related to certain other modal notions (such as being in a position to know, truth, or necessity). At times, the logic of justified belief is investigated in a purely

✉ Daniela Schuster
daniela.schuster@uni-konstanz.de

Leon Horsten
leon.horsten@uni-konstanz.de

[1] Universität Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany

propositional setting, at times it is investigated in a partially quantified setting (where one can quantify, for instance over objects or over reasons).[1]

Our aim in this article is to carry out a logical investigation of the concept of justified belief in a quantificational setting that is *highly expressive* in the specific sense that the objects of justified belief can explicitly be quantified over. Moreover, our investigation is *pure* in the sense that in our framework, we do not relate justified belief to other specific notions (such as truth or knowledge) or to other specific entities (such as reasons or agents). Indeed, we believe that it is beneficial to gain some clarity about the pure logical laws of justified belief before investigating the logical interaction between justified belief and other modal concepts.

The line of research to which we aim to contribute started around 1980 with (Thomason, 1980; Burge, 1978), building on earlier work such as (Hintikka, 1962; Kaplan & Montague, 1960; Myhill, 1960). It continued at a modest pace over the decades following that, up to the present day, see e.g. (Thomason, 1980; Koons, 1992; Cross, 2001; Cieśliński, 2017; Van Fraassen, 2022).

After a discussion of the background of the problem of the logic of justified belief (Sects. 2 and 3), we formulate a minimal base theory of principles and inference rules for justified belief, and develop a philosophical argument for the thesis that this base theory is incontrovertible (Sect. 4.1). Then we consider a list of additional principles concerning justified belief that have been advocated and criticised in the literature (4.2). We prove that almost every one of them cannot consistently be added to our base theory (4.3). Most of these inconsistency results are not our own: our aim in this section is mainly to collect, extend, and systematically connect results that are scattered in different places in the literature. We then proceed to discuss a few additional principles that can consistently be added to the base theory (Sect. 4.4). In the closing sections, we discuss philosophical arguments for and against the additional principles (Sects. 5 and 6), and sum up our findings (Sect. 7).

2 The framework

Justified belief is a relation between cognitive agents and entities of some kind. In this article, we keep the cognitive agent fixed, and we will call her Catrin. Moreover, we formally treat justified belief not as an operator but as a *predicate* (J). We do this because we want to work in an expressively strong framework. For instance, we want to be able to capture inferences such as:

$$\frac{\text{Catrin believes that fire is hot.}}{\text{There is something that Catrin believes.}}$$

In the epistemological literature, the objects of propositional attitudes such as belief are often taken to be propositions. From a formal perspective, this complicates matters, because there is currently no satisfactory theory of propositions. But the structural properties of propositions closely resemble those of sentences, and there are excellent

[1] See e.g. (Artemov, 2008; Heylen, 2020; Rosenkranz, 2018, 2021) or (Cieśliński, unpublished).

theories of the grammatical structure of sentences. So we will take the objects of the attitude of justified belief to be sentences.

It is well-known that, modulo coding, the grammatical structure of sentences is well described by elementary theories of arithmetic, such as Peano Arithmetic (PA). So we take Peano Arithmetic as a background theory of the formal theories of justified belief that will be considered. We will take mathematical induction to govern *all* properties of numbers. So our theories will be expressed in the language of arithmetic $\mathcal{L}_{\text{PA}}$ plus the new justified belief predicate J, and we call this extended language $\mathcal{L}_J$. And we take the background theory to be PA^J, which is Peano Arithmetic formulated in the extended language, with J allowed to occur in instances of the induction scheme. In the interest of readability, we will be somewhat sloppy in our notation, especially when it comes to the details of Gödel coding.

Although in our framework we are investigating justified belief only, it is rewarding to consider analogies between justified belief (by a fixed agent) on the one hand, and the concept of truth on the other hand. In formal truth theory, truth is treated as a predicate (T). If we work in the language $\mathcal{L}_{\text{PA}} + T$, which we may call $\mathcal{L}_T$, then we can generate the liar paradox from the unrestricted Tarski-biconditionals. This has profound implications for axiomatic theories of the fundamental laws governing truth see (Halbach, 1996). In epistemic contexts, similiar situations have occured, so that we find liar like paradoxes for epistemic notions, too.[2] Those epistemic analogues of the liar paradox will play a major role in the investigations that are carried out in this paper.

In epistemology, justified belief is often implicitly or explicitly treated as an operator instead of as a predicate. In the resulting expressively impoverished setting, liar-like, self-referential arguments cannot be carried out anymore. Despite this advantage, we consider this practice, for certain purposes, unsatisfactory, for it greatly diminishes our ability to identify fundamental laws governing justified belief. Epistemology, like other fundamental philosophical disciplines such as metaphysics, has *universal* ambitions: it aims to uncover basic principles that hold for *all* instances of justified belief (empirical beliefs, moral beliefs, mathematical beliefs, philosophical beliefs, etcetera). In particular, these basic principles should also cover beliefs about oneself and one's own beliefs.[3] In other words, it would be a dereliction of duty of the epistemologist to disregard self-referential beliefs from the outset. So, in order to guarantee sufficient expressiveness, we will take justified belief to be a predicate.

We take being *justified* in one's belief to be a matter of *having good reasons* for one's belief, where the reasons themselves have propositional structure. In this sense, we are working with an ontologically internalist conception of justification. We do not thereby exclude that there may be other, external ways of being epistemically warranted in one's belief.[4] For instance, suppose Catrin looks out of her window and on the basis of her perception forms the belief that there is an apple tree outside. Then we would normally say that this belief is epistemically warranted. Yet, the perception

[2] See e.g. (Horsten, 2011).

[3] Although in modal logic it is to some extent possible to talk about one's own belief via introspection principles, the combination of this with the possibility to quantify is in this setting not possible.

[4] One such theory can be found in Burge (1997).

on which it is based may not have propositional structure, and therefore not even being a candidate for being a *reason* for her belief. If that is so, then there are also warrants that do not take the form of reasons, and therefore—in the sense in which we use the term—are not justifications. Hence, we accept that there might be different types of epistemic warrants, but reserve the term justification for reason-based warrants. In this sense, for the purposes of this article, we want to stay neutral in the debate between internalism and externalism in epistemology. As we are talking about justified belief, we will be concerned with doxastic rather than so-called propositional justification (Rosenkranz, 2018, p. 310).

3 From Montague and Kaplan to Thomason

The aforementioned parallels between results in truth theory and similar results concerning epistemic notions led to a brace of interesting results in formal epistemology. Montague and Kaplan (and, independently, Myhill), proved a strengthening of Tarski's result about the undefinability of truth:

Theorem 1 (Kaplan and Montague (1960), Myhill (1960)) *Let S be the closure of* $\mathrm{PA}^T + \{T^\ulcorner\phi^\urcorner \rightarrow \phi \mid \phi \in \mathcal{L}_T\}$ *under the Necessitation rule*

$$\frac{\vdash \phi}{\vdash T^\ulcorner\phi^\urcorner}$$

then S is inconsistent.

Proof As for all the inconsistency results in this article, this is proved by a simple diagonal argument. The argument is based on a sentence α which is such that

$$1)\ \mathrm{PA}^T \vdash \alpha \leftrightarrow \neg T^\ulcorner\alpha^\urcorner,$$

which can be obtained by the Diagonal Lemma, see (Horsten, 2011, Theorem 12, p. 37). In order to derive the inconsistency assume

	2) $\neg\alpha$	*assumption*
	3) $T^\ulcorner\alpha^\urcorner$	*1), 2)*
	4) α	*3), T-Out axiom*
	5) $\bot$	*2), 4)*
6)	$\vdash \alpha$	*assumption-negation, 2)*
7)	$\vdash T^\ulcorner\alpha^\urcorner$	*6), Necessitation rule from S*
8)	$\vdash \neg T^\ulcorner\alpha^\urcorner$	*6), 1)*
9)	$\vdash \bot$	*7), 8)*

The Tarski-biconditionals (i.e., sentences of the form $T^\ulcorner\phi^\urcorner \leftrightarrow \phi$) are not valid when the predicate T is interpreted not as truth, but as 'it is known that'. But on the interpretation of T as 'it is known that', both the unrestricted schematic principle $T^\ulcorner\phi^\urcorner \rightarrow \phi$ (T-Out) and the unrestricted schematic rule

$$\frac{\vdash \phi}{\vdash T\ulcorner\phi\urcorner}$$

(Nec) seem plausible. So the argument for Theorem 1 yields an *epistemic paradox*. This is why, when the sentence α in Theorem 1 is given an epistemic interpretation, it is called the *knower sentence* in the literature, see e.g. (Kaplan & Montague, 1960; Anderson, 1983).

It is sometimes objected that this is only a paradox for a highly *idealised* notion of knowledge ("knowledge by God", perhaps). Whereas God's knowledge is (presumably) closed under Nec, we, as finite beings, can only ever apply Nec finitely many times. However, this argument is unconvincing. Only a few instances of Nec (only one, in fact) are required to obtain a contradiction. So Catrin can be assumed to have gone through this finite argument in S. This is exactly what we assume in the present article. So we have a paradox not for an idealised version of knowledge, but for ordinary knowledge by ordinary finite agents. We claim this not only for this specific theory S, but also for other inconsistent theories that we will consider later.

When we regard the Kaplan-Montague theorem as a theorem about truth, then there are two prima facie reasonable ways of reacting to it:

1. reject T-Out, but preserve Nec;
2. reject Nec, but preserve T-Out.

(Of course one can also reject both, and it is also possible to doubt the applicability of some of the laws of classical logic in this context.)

These two reactions have resulted in the two corresponding main families of axiomatic theories of truth:[5])

1. *FS*-like theories of truth;
2. *KF*-like theories of truth.

The same two prima facie reasonable reactions suggest themselves when we interpret the predicate T in theory S instead as knowledge. However, when we interpret T in the Kaplan-Montague theorem as justified belief, only the first reaction seems to have initial plausibility. Nec (for J) seems eminently plausible. Indeed, *proving* a sentence ϕ from basic logical principles concerning justified belief seems a paradigmatic way of justifying ϕ. (Indeed, the connection between justifying and proving is worth bearing in mind.) On the other hand, the justified belief-analogue of T-Out, let us call it J-Out, is totally implausible, as epistemology has taught us.

So the Kaplan-Montague theorem presents a paradox for knowledge, but not for justified belief. This raises the question whether the principles of S can somehow be weakened in such a way that a paradox for justified belief is obtained.

This question was first addressed by Thomason (1980). Consider the theory THO consisting of PA^J plus the following principles governing J:

1. $J\ulcorner\phi\urcorner \rightarrow J(\ulcorner J\ulcorner\phi\urcorner\urcorner)$;
2. $J(\ulcorner J\ulcorner\phi\urcorner \rightarrow \phi\urcorner)$;
3. $J\ulcorner\phi\urcorner$ if ϕ is a logical axiom;

[5] See (Horsten, 2011, chapter 8 and 9).

4. $J(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow (J\ulcorner\phi\urcorner \rightarrow J\ulcorner\psi\urcorner)$;
5. $\neg J\ulcorner\bot\urcorner$, where $\bot$ is your favourite contradiction.

Observe that J-Out does not belong to the axioms of THO, but a slight weakening of this axiom (Axiom 2) does. Observe also that the Rule 3 is a (weak) form of Nec. Then we have (Thomason (1980)):[6]

Theorem 2 *THO is inconsistent.*

Thomason describes his theorem as a paradox for *idealised belief* (Thomason, 1980, p. 392). But belief is a logically very unconstrained notion: a madman may believe or not believe just about anything. It is more accurate to describe it as a paradox about *rationally justifiable belief* (Koons, 1992, p. 14), or even ordinary *justified* belief. As with the Kaplan-Montague paradox for knowledge, Thomason's argument indeed yields a paradox about *ordinary* justified belief, since, in the light of what we said on p. 4, we may (and will) assume that Catrin has gone through the handful of logical steps that it takes to reach the contradiction, without assuming logical omniscience.[7]

4 Inconsistency, consistency, strength

Friedman and Sheard (1987) step back and adopt a more general perspective to truth and paradox. They start with a base theory containing PA^T and a few basic principles of truth that at least for the purposes of their discussion can be taken to be non-negotiable. Then they draw up a list L of principles that can make some prima facie claim to being basic truth laws. Subsequently they show that certain subsets of L are inconsistent over the base theory, and that certain other subsets of L are consistent. Moreover, they give a proof-theoretic analysis of some of these consistent truth systems.

In this section, we want to at least make a start at providing a similar overview for justified belief and thereby helping to explicate the concept of justified belief and the interaction of certain prima facie plausible principles of justified belief.

4.1 Base theory

We take it that PA^J must be included in the base theory, which we will call B. The reason is that the principles of classical logic hold not only for arithmetical formulas, but also for formulas including the justification predicate J. Moreover, the induction axiom is open-ended, so it also holds for formulas that include J. Let us now turn to the non-arithmetical axioms of B.

First, *epistemic closure* (CL) is the principle

$$(J\ulcorner\phi\urcorner \wedge J(\ulcorner\phi \rightarrow \psi\urcorner)) \rightarrow J\ulcorner\psi\urcorner.$$

[6] We do not here go through the simple diagonal argument that establishes Thomason's theorem.

[7] Cross (2001) also argues that idealisation plays no role in the paradox. Of course, it is important to note that any logical system comes with a certain degree of idealisation. We thank an anonymous reviewer for this remark.

This principle is also included in B. It expresses that logical reasoning preserves justification.

Second, *epistemic coherence* (CO) is the principle $\neg J^{\ulcorner}\bot^{\urcorner}$, where $\bot$ is your favourite explicit contradiction of the form $p \wedge \neg p$. If Catrin's reasoning process has reached such a conclusion, then something has gone wrong: she is not justified in believing this conclusion. We include CO in B. Together with CL, this principle also prevents Catrin from being justified in principles from which a contradiction follows by logical reasoning.

Third, we have argued above (p. 4), that the *Necessitation* rule (Nec) should be accepted. It expresses the thought that proof (from the laws and rules of justification plus the principles of arithmetic) yields justification. We also include it in B.

In sum, we take our Base Theory B to be

$$\mathrm{PA}^{\mathrm{J}} + \mathrm{CL} + \mathrm{CO} + \mathrm{Nec}.$$

An agent who satisfies *all* theorems of B must have justified belief in infinitely many statements, and is for this reason super-human. But for every *finite* subset of the theorems of B, there is a finite rational agent whose justified beliefs are exactly the statements belonging to this finite set. So the base system B is best seen as describing not one infinite ideal rational agent, but an infinite family of finite non-idealised rational agents: an infinite family of finite "Catrins". One can go even further, and hold that only those small finite sets of theorems of B that have short proofs in B are intended to describe flesh and blood rational human agents.

In view of this and of what we said on p. 4, we stress that Nec and epistemic closure do not commit us to Catrin being an "unrealistically idealised" epistemic agent: the proofs in B and its putative extensions that will be considered are all short, and contain very few instances of CL and Nec. In fact all proofs require at most five applications of either Nec or CL.[8] This is a completely realistic job for a non-ideal agent.

Still, whether our choice of B as non-negotiable base theory is a judicious one, is a non-trivial question.[9]

Epistemic closure is a controversial principle. As Smith[10] in a recent article rightly points out (Smith, 2018, Section 1), some swear by it,[11] others reject it. Doubts about CL relate to *lottery paradox*-like situations. Suppose I give a party for 10 people, and for each of them I justifiedly believe, with a degree of certainty of 0.9, that (s)he will come to the party. Moreover, suppose that I know that these 10 people will arrive at their ultimate decision whether to attend completely independently. Then, by CL, I can come justifiedly to believe that they will all attend. But at the same time, my credence that they all attend will be < 0.4. Our response to this objection is that in the scenario, I was not justified in believing, for each of these 10 people, that (s)he will

[8] Or the principle of *Agglomeration*, which follows from both: see below.

[9] Likewise, it is a non-trivial question whether Friedman and Sheard chose the right base theory for truth in Friedman and Sheard (1987).

[10] The example of Smith rather uses the agglomeration principle (Agg) than epistemic closure. We will show at the send of this section, though, that the two principles are related.

[11] For instance (Cieslinski, 2017, p. 254), (Artemov, 2008, p. 482). In (Kuznets, 2008, pp. 35–36) common modal justificiation logics are considered which do all include the closure principle.

attend. This is because the notion of *belief* in the concept of justified belief that we are investigating, is one of *unreserved*, *unqualified* belief. Given this concept of justified belief, I was, for each possible attendant, only justified to have a *graded* belief that (s)he will attend.

Another critical view on epistemic closure can be found in the discussion of Heylen (2016) and Rosenkranz (2016). Heylen discusses the notion of *being in a position to know* and introduces a counterexample to the agglomeration principle ($K\phi \wedge K\psi \rightarrow K(\phi \wedge \psi)$) for this notion.[12] Since agglomeration follows from closure (assuming the subject is in a position to know some easy tautology), closure has to be rejected. Rosenkranz (2016) picks up on this argument and continues to show that closure has to be rejected for justification, too. This, though, heavily relies on Rosenkranz's principle $K\phi \rightarrow J\phi$, see also (Rosenkranz, 2018) and (Rosenkranz, 2021). This argument is not applicable for our predicate, though. Since we are considering J to be *justified belief*, or what Rosenkranz (2018) calls *doxastic justification*[13] rather than propositional justification, it is not that case that *having* a justified belief in ϕ can be derived from being in a position to know ϕ, which makes the argument against the closure for J not applicable to our theory.

Epistemic coherence is, in comparison with epistemic closure, less controversial. *Pace* dialethism, very few philosophers believe that one can ever be justified in believing a blatant contradiction.[14] Still, in Sect. 7 we briefly discuss the ramifications of excluding epistemic coherence from the base theory for the inconsistency results that will be presented in Sect. 4.3.

Lastly, there is Nec. As said on p. 4 and p. 5, we subscribe to this inference rule.[15] We take the basic logical principles of justified belief to express elementary conceptual truths about justified belief. So reflection on the content of these principles yields justified belief in them. However, in the light of what was said earlier (p. 4), we should not expect Nec to be automatically extended when Catrin's theory[16] is extended.

We want to note at this point that our base theory also proves the principle of agglomeration (Agg), $(J\ulcorner\phi\urcorner \wedge J\ulcorner\psi\urcorner) \rightarrow J(\ulcorner\phi \wedge \psi\urcorner)$, which we will use in the upcoming section:[17]

1)	$\vdash \phi \rightarrow (\psi \rightarrow (\phi \wedge \psi))$	*tautology*
2)	$\vdash J(\ulcorner\phi \rightarrow (\psi \rightarrow (\phi \wedge \psi))\urcorner)$	*Nec, 1)*
	3) $J\ulcorner\phi\urcorner \wedge J\ulcorner\psi\urcorner$	*assumption*
	4) $J\ulcorner\phi\urcorner$	*3), conjunction elimination*
	5) $J(\ulcorner\psi \rightarrow (\phi \wedge \psi)\urcorner)$	*CL, 2), 4)*
	6) $J\ulcorner\psi\urcorner$	*3), conjunction elimination*
	7) $J(\ulcorner\phi \wedge \psi\urcorner)$	*CL, 5), 6)*

[12] In the counterexample ψ is the sentence "no one knows that ϕ" and ϕ itself is a sentence that is easily verifiable but of little interest, such as "there is an even number of books in my office", (Heylen, 2016).

[13] Note that in (Rosenkranz, 2021, p. 107), Rosenkranz uses a different notion of doxastic justification that does not require having a belief.

[14] See for instance (Rosenkranz, 2018, p. 313), (Heylen, 2020), (Cieslinski, 2017, p. 254).

[15] Cieslinski also subscribes to this rule: see (Cieslinski, 2017, p. 254).

[16] By Catrin's theory we mean the set of sentences she believes.

[17] Thanks to an anonymous referee for urging us to spell out this argument.

8) $\vdash (J\ulcorner\phi\urcorner \wedge J\ulcorner\psi\urcorner) \rightarrow J(\ulcorner\phi \wedge \psi\urcorner)$ *implication introduction*

4.2 Candidate principles for justified belief

We will investigate the following additional principles for justified belief:

TH	*Thomason's principle:*	$J(\ulcorner J\ulcorner\phi\urcorner \rightarrow \phi\urcorner)$
GH	*Gödel-Hilbert principle:*	$J(\ulcorner\phi \rightarrow J\ulcorner\phi\urcorner\urcorner)$
PI	*Positive introspection:*	$J\ulcorner\phi\urcorner \rightarrow J(\ulcorner J\ulcorner\phi\urcorner\urcorner)$
CPI	*Converse positive introspection:*	$J(\ulcorner J\ulcorner\phi\urcorner\urcorner) \rightarrow J\ulcorner\phi\urcorner$
NI	*Negative introspection:*	$\neg J\ulcorner\phi\urcorner \rightarrow J(\ulcorner\neg J\ulcorner\phi\urcorner\urcorner)$
CNI	*Converse negative introspection:*	$J(\ulcorner\neg J\ulcorner\phi\urcorner\urcorner) \rightarrow \neg J\ulcorner\phi\urcorner$
MO	*Moore's principle:*	$\neg J(\ulcorner\phi \wedge \neg J\ulcorner\phi\urcorner\urcorner)$
CI	*Cieslinski's principle:*	$J\ulcorner\forall x J\ulcorner\phi(x)\urcorner\urcorner \rightarrow J\ulcorner\forall x\phi(x)\urcorner$
WRef	*Weak Reflection:*	$\forall x : Bew_{\mathrm{PA}^J}(x) \rightarrow J\ulcorner x\urcorner$

TH was adopted in (Thomason, 1980, p. 391). For versions of the principle **GH**, see (Boolos, 1995, p. 294) and (Wang, 1974, p. 324–325). **PI**, **CPI**, **CNI** and **NI** are discussed widely in contemporary epistemology; early discussions investigations of these principles can be found in (Hintikka, 1962) and (Lenzen, 1980, p. 38–39). **MO** is discussed in (Van Fraassen, 2022, p. 17) and (Cross, 2001). **CI** and **WRef** are discussed in (Cieśliński, 2017) and (Cieśliński, unpublished). In their discussion of these principles, most of these scholars interpret J not as justified belief but as some related notion; a philosophical discussion of many of these principles in terms of justification can be found in (Rosenkranz, 2018).

The introspection principles listed above have the form of a material implication. Certain *rule forms* of such principles might be considered. For instance, one might wonder about the following rule-form of converse postive introspection:

$$\frac{\vdash J(\ulcorner J\ulcorner\phi\urcorner\urcorner)}{\vdash J\ulcorner\phi\urcorner},$$

or even, more generally, about the converse of Nec, namely the Co-Necessitation rule CoNec:

$$\frac{\vdash J\ulcorner\phi\urcorner}{\vdash \phi}.$$

Although the rules versions can be derived from the material implication versions, this is not the case the other way around. Investigating only the rule versions would, thus, lead to weaker theories. These matters are left for another occasion. We merely pause to note that the philosophical justification of CoNec is no trivial matter, given that justified belief is not a factive notion.

4.3 Some impossibility theorems

It turns out that many of the additional principles in Sect. 4.2 are inconsistent with the base theory B. Indeed, simple diagonal arguments show:

Theorem 3 (A) *When the base theory B is extended by any of the following as extra axioms, an inconsistent theory results:*

(a) ***PI + TH****;*
(b) ***TH****;*
(c) ***GH****;*
(d) ***PI****;*
(e) ***NI****.*

(B) *When the theory theory $B - CO$ (i.e., the system B without epistemic coherence) is extended by any of the following as extra axioms, an inconsistent theory results:*

(f) ***CNI****;*
(g) ***MO****.*

Proof a) has already been proven by Thomason (1980).

b)

1)	$\vdash \alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner$	*diagonal lemma*
2)	$\vdash J(\ulcorner\alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner\urcorner)$	*Nec, 1)*
3)	$\vdash J(\ulcorner J\ulcorner\alpha\urcorner \rightarrow \alpha\urcorner)$	***TH***
4)	$\vdash J(\ulcorner(J\ulcorner\alpha\urcorner \rightarrow \alpha) \wedge (\neg J\ulcorner\alpha\urcorner \rightarrow \alpha)\urcorner)$	*Agg, 2), 3)*
5)	$\vdash ((J\ulcorner\alpha\urcorner \rightarrow \alpha) \wedge (\neg J\ulcorner\alpha\urcorner \rightarrow \alpha)) \rightarrow \alpha$	*tautology, tertium non datur*
6)	$\vdash J(\ulcorner((J\ulcorner\alpha\urcorner \rightarrow \alpha) \wedge (\neg J\ulcorner\alpha\urcorner \rightarrow \alpha)) \rightarrow \alpha\urcorner)$	*Nec, 5)*
7)	$\vdash J\ulcorner\alpha\urcorner$	*CL, 4), 6)*
8)	$\vdash \neg\alpha$	*1), 7) (contraposition)*
9)	$\vdash J\ulcorner\neg\alpha\urcorner$	*Nec, 8)*
10)	$\vdash J\ulcorner\bot\urcorner$	*Agg, 7), 9)*
11)	$\vdash \bot$	*CO, 10)*

c)

1)	$\vdash \alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner$	*diagonal lemma*
2)	$\vdash J(\ulcorner\alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner\urcorner)$	*Nec, 1)*
3)	$\vdash J(\ulcorner\alpha \rightarrow J\ulcorner\alpha\urcorner\urcorner)$	***GH***
	4) $J\ulcorner\alpha\urcorner$	*assumption*
	5) $J(\ulcorner J\ulcorner\alpha\urcorner\urcorner)$	*CL, 3), 4)*
	6) $J(\ulcorner\neg J\ulcorner\alpha\urcorner\urcorner)$	*CL, 2), 4)*
	7) $J\ulcorner\bot\urcorner$	*Agg, 5), 6)*
	8) $\bot$	*CO, 7)*
9)	$\vdash \neg J\ulcorner\alpha\urcorner$	*assumption-negation, 4)*
10)	$\vdash \alpha$	*1), 9)*

11) $\vdash J^{\ulcorner}\alpha^{\urcorner}$	*Nec, 10)*
12) $\vdash \bot$	*9), 11)*

d)

1) $\vdash \alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner}$	*diagonal lemma*
2) $\vdash J(^{\ulcorner}\alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*Nec, 1)*
3) $\vdash J^{\ulcorner}\alpha^{\urcorner} \rightarrow J(^{\ulcorner}J^{\ulcorner}\alpha^{\urcorner\urcorner})$	***PI***
4) $\neg\alpha$	*assumption*
5) $J^{\ulcorner}\alpha^{\urcorner}$	*1), 4) (contraposition)*
6) $J(^{\ulcorner}J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*3), 5)*
7) $J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*CL, 2), 5)*
8) $J^{\ulcorner}\bot^{\urcorner}$	*Agg, 6), 7)*
9) $\bot$	*CO, 8)*
10) $\vdash \alpha$	*assumption-negation, 4)*
11) $\vdash J^{\ulcorner}\alpha^{\urcorner}$	*Nec, 10)*
12) $\vdash \neg J^{\ulcorner}\alpha^{\urcorner}$	*1), 10)*
13) $\vdash \bot$	*11), 12)*

e)

1) $\vdash \alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner}$	*diagonal lemma*
2) $\vdash J(^{\ulcorner}\alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*Nec, 1)*
3) $\vdash \neg J^{\ulcorner}\alpha^{\urcorner} \rightarrow J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	***NI***
4) $\neg J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*assumption*
5) $J^{\ulcorner}\alpha^{\urcorner}$	*3), 4) (contraposition)*
6) $J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*CL, 2), 5)*
7) $\bot$	*4), 6)*
8) $\vdash J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*assumption-negation, 4)*
9) $\vdash J^{\ulcorner}\alpha^{\urcorner}$	*CL, 2), 8)*
10) $\vdash \neg\alpha$	*1), 9)*
11) $\vdash J^{\ulcorner}\neg\alpha^{\urcorner}$	*Nec, 10)*
12) $\vdash J^{\ulcorner}\bot^{\urcorner}$	*Agg, 9), 11)*
13) $\vdash \bot$	*CO, 12)*

f)

1) $\vdash \alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner}$	*diagonal lemma*
2) $\vdash J(^{\ulcorner}\alpha \leftrightarrow \neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*Nec, 1)*
3) $\vdash J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner}) \rightarrow \neg J^{\ulcorner}\alpha^{\urcorner}$	***CNI***
4) $\neg\alpha$	*assumption*
5) $J^{\ulcorner}\alpha^{\urcorner}$	*1), 4) (contraposition)*
6) $J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*CL, 2), 5)*
7) $\neg J(^{\ulcorner}\neg J^{\ulcorner}\alpha^{\urcorner\urcorner})$	*5), 3) (contraposition)*
8) $\bot$	*6), 7)*
9) $\vdash \alpha$	*assumption-negation, 4)*

10) $\vdash J\ulcorner\alpha\urcorner$	*Nec, 9)*
11) $\vdash \neg J\ulcorner\alpha\urcorner$	*1), 9)*
12) $\vdash \bot$	*10), 11)*

g)

1) $\vdash \alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner$	*diagonal lemma*
2) $\vdash J(\ulcorner\alpha \leftrightarrow \neg J\ulcorner\alpha\urcorner\urcorner)$	*Nec, 1)*
3) $\vdash \neg J(\ulcorner\alpha \wedge \neg J\ulcorner\alpha\urcorner\urcorner)$	***MO***
4) $J\ulcorner\alpha\urcorner$	*assumption*
5) $J(\ulcorner\neg J\ulcorner\alpha\urcorner\urcorner)$	*CL, 2), 4)*
6) $J(\ulcorner\alpha \wedge \neg J\ulcorner\alpha\urcorner\urcorner)$	*Agg, 4), 5)*
7) $\bot$	*3), 6)*
8) $\vdash \neg J\ulcorner\alpha\urcorner$	*assumption-negation, 4)*
9) $\vdash \alpha$	*1), 8)*
10) $\vdash J\ulcorner\alpha\urcorner$	*Nec, 9)*
11) $\vdash \bot$	*8), 10)*

Observe that every single one of these inconsistency results follow by straightforward application of the axioms and rules of B to the "knower" sentence. Several of these results can already be found in the literature. Case a) is Thomason's theorem. Case b) is a slight improvement on Thomason's theorem, so is case d), as both b) and d) only use one of the two additional principles that were used in Thomason's theorem a). Moreover, Koons observes that Gödel's incompleteness theorem can be translated into epistemic logic and is then basically represented by case d) (Koons, 1992, p. 54–55). Burge, on the other hand, used both the positive and negative introspection to prove an inconsistency (Burge, 1978, p. 29;) this is also discussed in (Koons, 1992, p. 53). Cases d) and e) are likewise slight strengthenings of this result. Some of the remaining inconsistency results are undoubtedly folklore; we just have not found references to them in the literature.[18]

Theorem 3 reveals that all candidate additional principles are straightforwardly inconsistent over B, except **CPI**, **CI** and **WRef**. Moreover, most of the inconsistency results rely on Nec. So Nec imposes strong restrictions on the logic of justification.

4.4 Quantification

The base system B does not say anything about the logical interaction between the universal quantifier and the modal predicate J. The collective scholarly experience in axiomatic truth theory suggests that such a system is likely to be conservative over the background theory. Indeed, for B this is the case:

Theorem 4 *B is arithmetically conservative over* PA.

Proof This follows from Theorem 3 of (Cieśliński et al., 2022). The translation τ which distributes over logical connectives, is homophonic for purely mathematical

[18] Similar results can also be found in Égré (2005).

statements, and which is such that $\tau(J\phi)$ is set equal to $\Pr(\tau(\phi)) = 1$ translates theorems of B into theorems of the system RKf, which is (by Theorem 3 of (Cieśliński et al., 2022)) proof-theoretically conservative over its background theory.[19] □

Cieśliński has developed a formal theory of 'believability' (Cieśliński, 2017), which is a notion which is closely related to that of justified belief. A few years after the publication of his book Cieśliński proposed an improved axiomatic theory of believability (Cieśliński, unpublished): we base our discussion on the later version of his theory.

Cieśliński's original theory contains a rule form of **CI**; in Cieśliński's later theory, **CI** is one of the axioms. This axiom contains information about the interaction between universal quantification and J, and is a strengthening of converse positive introspection (CPI). Indeed, **CI** expresses a sort of effective, finitised ω-rule for J in the following sense. Hilbert's ω-rule says that from an infinite list of premises $\phi(0)$, $\phi(1)$ we can conclude $\forall n\phi(n)$. Since this rule has infinitely many premises, it cannot humanly be applied and is, thus, not effective. However, the antecedent of **CI** requires that we have *one uniform justification* of the fact that $\forall n J\Phi(n)$. This finitises the rule and makes it suitable as a principle of reasoning.

Observe that we did not prove inconsistency results for **CI**. Cieśliński proved that it is consistent not only to add **CI**, but in addition also **WRef** to the base system:

Theorem 5 $B + \mathrm{WRef} + \mathrm{CI}$ *is arithmetically sound.*

Proof This follows from Cieśliński (unpublished), Theorem 5 □

This immediately entails the following corollary:

Theorem 6 $B + \mathrm{WRef} + \mathrm{CPI}$ *is arithmetically sound.*

Proof This follows immediately from Theorem 5, since **CPI** follows from **CI**. □

Since **CI** has the flavour of a constructive ω-rule, we might expect that it adds some proof theoretic strength. In Cieśliński's work, a slight weakening of the system $B+CI$ plays a central role, namely $B + \mathrm{WRef} + \mathrm{CI} - \mathrm{CO}$, which he calls $Bel^*(\mathrm{PA})$.[20] He shows that this system is significantly non-conservative over PA.

We recall the definition of *Uniform Reflection* extensions of a ground theory S:

Definition 1 $- URF^0(S) \equiv S$;
$- URF^{n+1}(S) \equiv URF^n + \{\forall x : Bew_{URF^n}(\phi(x)) \rightarrow \phi(x) \mid \phi(x) \in \mathcal{L}_J\}$.

Furthermore, we define the *internal logic* of a logical system of justified belief S (denoted as $Int(S)$) as those sentences ϕ such that $J\phi \in S$. Since J is not factive, $Int(S)$, rather than S, should be taken as showing the strength of S. There will be theorems that are provable using **CI** and that are of the form $J\ulcorner\phi\urcorner$ with ϕ being an arithmetical statement that is unprovable in the background theory PA. Because we do not have the principle J-Out, this will not result in outright arithmetical non-conservativness of the theory of S.

More precisely, Cieśliński shows:

[19] Łełyk (unpublished) proved an even stronger result, namely, that the theory FS^- is truth-definable in the theory consisting of the uniform Tarski-biconditionals added to the background mathematical theory.

[20] See Cieśliński (unpublished).

Theorem 7 *For every natural number n:*

$$URF^n(B) \subseteq Int(Bel^*(\mathrm{PA})).$$

Proof See (Cieslinski, unpublished, Theorem 8). Cieśliński proves that for every natural number n and for all $\phi(x) \in L_{PAB}$: $B\forall(xPr_{URF^n(S)}(\phi(x)) \rightarrow x) \in Bel^*(PA)$. By the definition of the concept of internal logic above, this means that for all n and for all $\phi(x) \in L_{PAB}$: $\forall(xPr_{URF^n(S)}(\phi(x)) \rightarrow x) \in Int(Bel^*(PA))$. □

The positive news here is of course that we have finally found some additional principles that are consistent (and indeed arithmetically sound) over our base theory. The fact that **CI** is mainly responsible for considerable extra arithmetical strength may be taken as a reason not to take it as belonging to the *basic* laws of justified belief. Something similar might be said about **WRef**, which explicitly has the form of a proof theoretic reflection principle.

5 Discussion of the additional principles

The additional principles that were presented in Sect. 4.3 have been discussed in the epistemological literature, and different arguments for and against them have been formulated. However, often the arguments are embedded in a discussion where J is taken to be not justified belief, but another epistemic notion, such as some other kind of belief, knowledge, subjective provability, etcetera. So we cannot simply transfer all epistemological arguments from the literature to the notion of justified belief that we are discussing in this article.

5.1 Thomason and Gödel–Hilbert

Koons (1992, p. 49) argues for Thomason's original principle **TH** but interprets J as subjective provability. He argues that every instance of the formula $J\ulcorner\phi\urcorner \rightarrow \phi$ is self-evident, and hence also subjectively provable. This argument might (or might not) be plausible when taking J to represent subjective provability. But it does not carry over to justified belief. After all, provability (in its informal sense) is a factive notion, whereas justified belief is not. Nonetheless, one might think that despite the fact that Catrin rationally believes or even knows that some of her justified beliefs are false, it is rational for her to believe *for each individual* ϕ that $J\ulcorner\phi\urcorner \rightarrow \phi$, on the basis that rational belief is a good (but not infallible) guide to truth, i.e., one might believe that **TH** is still true.

The principle **GH** is a generalisation of a thesis of Kant, who asserted that (Kant, 1993, A 476/B 504):

> […] there are sciences [and pure mathematics is one of them (Kant, 1993, A 480/B 508)] the very nature of which requires that every question arising within their domain should be completely answerable in terms of what is known, inasmuch as the answer must issue from the same sources from which the question proceeds.

In other words, by asking unanswerable questions while asserting that only reason can answer them, reason would be irrational. Hence, when asking the question whether ϕ, reason should be able to answer that question, i.e., decide whether $J\ulcorner\phi\urcorner$ or $J\ulcorner\neg\phi\urcorner$. With this we get the principle $J(\ulcorner J\ulcorner\phi\urcorner \vee J\ulcorner\neg\phi\urcorner\urcorner)$. This principle is a consequence of the **GH** principle.

His own incompleteness theorems notwithstanding, some such belief seems also to have been held by Gödel. In any case, **GH** is an expression of a form of *rational optimism* (Wang, 1974, p. 325), to which also Hilbert gave voice in his famous slogan that "in mathematics, there can be no *ignoramibus*."

On the other (pessimistic) side, we may have good reasons to know, for certain specific propositions, that they are *absolutely undecidable*. Some believe that even in arithmetic some such statements can be found: Feferman and Solovay speculate that the question "Is it true that if w is the sequence of the first $2^{2^{100}}$ terms in the binary expression of $7\pi - 3$, then the last term of w is 0?" may be a case in point (Feferman & Solovay, 1990, p. 292).

5.2 Moore and introspection

For Moore's principle the situation is more positive. Already Hintikka (1962, p. 125) defends **MO** for the notion of belief. Explicit arguments for **MO** for the notion of justified belief can be found in (Smithies, 2012). Smithies argues that instances of Moore's paradox can be found not only for belief, but also for knowledge and justification. In the same way that it is somehow wrong for someone to assert "p, but I do not believe that p", it seems irrational for someone to assert "p, but I do not have a justification for believing p" or "I have a justification for believing p but it is not the case that p", (see Smithies, 2012, pp. 283–284).

Moreover, Smithies argues that there is a tension between *accepting* **MO** on the one hand, and *rejecting* positive or negative reflection principles on the other hand. In this argument, he uses what he calls an *exhaustiveness* assumption: for each sentence ϕ, I am either justified in believing ϕ, justified in disbelieving ϕ (i.e., justified in believing $\neg\phi$), or justified in believing that I withhold judgement (W) about ϕ. He shows that if any of the four introspection principles fail, and the exhaustiveness assumption holds, then **MO** must fail, too. For example in the case of **PI**: If one takes **PI** to be false, then there is a sentence ϕ, such that $J\ulcorner\phi\urcorner$ and $\neg J(\ulcorner J\ulcorner\phi\urcorner\urcorner)$. From $\neg J(\ulcorner J\ulcorner\phi\urcorner\urcorner)$ Smithies concludes (using exhaustiveness) that I am either justified in disbelieving that I am justified in ϕ or I am justified in believing that I withhold whether I am justified in ϕ, so we have either $J\ulcorner\phi\urcorner \wedge J(\ulcorner\neg J\ulcorner\phi\urcorner\urcorner)$ or $J\ulcorner\phi\urcorner \wedge J(\ulcorner W(\ulcorner J\ulcorner\phi\urcorner\urcorner)\urcorner)$. In either case, using the agglomeration principle, we obtain a contradiction with **MO**. Since, for Smithies, **MO** is non-negotiable, he also accepts all four introspection principles for justified belief (Smithies, 2012, p. 286). A similar reductio can be distilled for **NI**, **CNI** and **CPI** (see Smithies, 2012, pp. 285–287).[21]

[21] Note that the four introspection principles are not independent. In our base theory, using CO and Agg, you can derive **CNI** from **PI**. For this, assume **CNI** to be false, so $J(\ulcorner\neg J\ulcorner\phi\urcorner\urcorner) \wedge J\ulcorner\phi\urcorner$. From $J\ulcorner\phi\urcorner$ you get $J(\ulcorner J\ulcorner\phi\urcorner\urcorner)$ with **PI**. With this and the first conjunct from the conjunction, you get $J\ulcorner\bot\urcorner$ using Agg. This is a contradition to CO. Similarly, you can derive **CPI** from **NI**.

More considerations about relevant principles for the notion of justification can be found in (Rosenkranz, 2018).[22] In his paper, Rosenkranz considers an all-things considered notion of justification which aims at being neutral on the debate about externalism versus internalism and a knowledge-first approach. The presented minimal logic of this neutral account of justification includes necessitation but neither epistemic closure, nor epistemic coherence. Rosenkranz's notion of justification is different from ours, as he considers only propositional justification which does not entail having a belief, in contrast to our notion of justified belief, which can be described as doxastic justification.[23] Although epistemic closure fails for J in his theory, still he introduces a rule for deducing $\vdash K\phi \rightarrow K\psi$ from $\vdash \phi \rightarrow \psi$, which in combination with Nec (which Rosenkranz accepts in (Rosenkranz, 2018)) comes quite close to closure. However, in (Rosenkranz, 2021), Rosenkranz presents a more articulated story, describing a logic for justification that does not use Nec, stating that this rule is "way too strong" and "highly suspect" (Rosenkranz, 2021, pp. 86 and 186). In his theory from 2018, though, Rosenkranz proves (by the use of others principles connecting justification to the other introduced modality of "being in a position to know") Moore's principle, and asserts that this principle is "eminently plausible" (Rosenkranz, 2018, p. 324).

Moreover, Rosenkranz considers an extension for his logical background theory. He calls this extension a logic of *luminous justification*; he takes it to be somewhat more internalism-directed but still compatible with many externalistic views. In this extension, Rosenkranz (2018, pp. 325–327) proves the third principle of our base theory, namely epistemic coherence.[24] Moreover, within this logic the principles of (converse) positive introspection and (converse) negative introspection are all four provable for justification.

Although the extended logic, from which these four principles are obtained, might slightly favor internalism, Rosenkranz still argues for their plausibility and stresses that they are obtained from externalist-friendly principles; see (Rosenkranz, 2018, p. 327). Moreover, Rosenkranz' theory strongly suggests the plausibility of Moore's principle, since this principle was already obtained in his minimal logic.[25]

These considerations are in harmony with conclusions that are reached in (Van Fraassen, 2019). Van Fraassen investigates a notion of belief that is stipulatively taken to be *self-transparent*. Then he argues, much like Smithies, that Moore's principle should hold, and constructs a class of models in which these constraints are all satisfied. Thus van Fraassen includes **PI** and **CPI** into his logic of a self-transparent believer, (Van Fraassen, 2019, 2022), as his modelled believer cannot be wrong about his own beliefs and is aware of them. Moreover, Van Fraassen (2022, p. 17) derives Moore's principle for his logic, using **PI** and **CPI**.

[22] For a critique of Rosenkranz' logic of justification, see Heylen (2020).

[23] As we stated in footnote 13, in (Rosenkranz, 2021, p. 107), Rosenkranz uses the term doxastic justification as a *sui generis* notion that does not require having a belief.

[24] In fact, Rosenkranz proves his principle D_J: $J\neg\phi \rightarrow \neg J\phi$. From this principle, though, CO can be derived easily in our and in Rosenkranz's logic.

[25] Note again that not even epistemic coherence, which is a member of our base theory, is obtained in his minimal logic.

All this does not mean that introspection principles are uncontroversial in epistemology. Firstly, it is not clear that all four introspection principles that we have considered stand or fall together.[26] Secondly, and perhaps more importantly, over the past decades all introspection principles have come to be regarded with suspicion. This is, to a significant extent, a consequence of Williamson's attack against *luminosity* theses in epistemology (Williamson, 2000, chapter 4). On this view, mental states are *in general* not fully transparent to those who have them. So *all* introspection principles for the notion of justified belief are questionable. For instance, Williamson argues against a form of positive introspection for even in mathematical cases, stating that one could very well have a proof of a theorem without being in a position to know that one has a proof (Williamson, 2000, p. 111).[27]

6 The restrictiveness of the minimal theory of justified belief

One of the principal aims of epistemology is to uncover *absolutely general* laws governing doxastic concepts such a justified belief. We have argued that the axioms and rules of the minimal basic system B may qualify as such. What about the additional principles that were introduced in Sect. 4.2?

Theorem 3 shows that many of them cannot be consistently added to B. This means that they fail to qualify as theorems governing justified belief. This implies that we can *bypass the philosophical arguments* in the literature that support them: these arguments miss their mark. Nonetheless, these arguments need not be without merit. It is possible that some authors who defend these principles, *implicitly* assume a *typing restriction* on the version of the principle that they have in mind, whereby self-referential counterexamples such as the knower sentence are ruled out. Of course, where this is so, it would be helpful if any restriction on the principles defended were explicitly stated.

What about the quantificational principle **CI** and its consequence **CPI**? We have seen that they can be consistently added to B (Theorem 5). Nevertheless, we think that **CPI**, and therefore also **CI**, is questionable. There seem to be cases where Catrin is justified in believing that she justifiably believes a proposition ϕ, without being actually justified in believing ϕ. Suppose that Catrin finds an apparent proof, which is complicated and long, for a mathematical statement ϕ. She takes her argument to be a valid mathematical proof, and checks it several times. She lets some of her colleagues check her proof, too: no one finds a mistake. Yet there is a subtle mistake in Catrin's mathematical argument. In this situation, it seems that Catrin is not justified in ϕ, since her argument contains a mistake. Nonetheless she is justified in believing that she is justified in believing ϕ. She fulfilled her epistemic obligations in that regard:

[26] Lenzen (1980, pp. 66–70), for example, argues for positive introspection but against negative introspection for knowledge. However, he accepts (converse) positive *and* negative introspection for belief: (Lenzen, 1980, pp. 38–39). Hintikka (1962, p. 123) argues for positive introspection for belief, but rejects converse positive introspection.

[27] However, in very recent work Rosenkranz (2021, pp. 76–79) argues against Williamson's anti-luminosity for instance for the condition $\neg K(\ulcorner \neg K \ulcorner \phi \urcorner \urcorner)$, which is equivalent to $J \ulcorner \phi \urcorner$ for Rosenkranz. The luminousity of $\neg K(\ulcorner \neg K \ulcorner \phi \urcorner \urcorner)$ and Rosenkranz' argument for it are questioned again in (Smith, 2022).

she did all she could do to secure her belief that she has justified the conclusion of her mathematical argument.

In sum, already the minimal system B is *highly restrictive*. Moreover, among the class of statements that can be consistently added, it is not easy to find ones that are plausible. Of course we have mainly looked at propositional modal principles, **CI** being the only exception. There may well be quantificational principles extending B that are promising, but this exceeds the scope of the present investigation.

It may be of interest to note that certain results are deducible even without the principle of epistemic coherence CO. Let the theory B without CO be called B-CO. In Sect. 4.3 we saw that adding **CNI** (f) or adding **MO** (g) to the base theory leads to inconsistency even without CO. Moreover, the remaining principles lead to more fine-grained results when CO is left out of the base theory. We found that B-CO + **CNI** $\vdash \bot$, B-CO + **MO** $\vdash \bot$, and we easily see that B-CO + **TH** $\vdash J\ulcorner\bot\urcorner$, B-CO + **NI** $\vdash J\ulcorner\bot\urcorner$, B-CO + **GH** $\vdash J\ulcorner\bot\urcorner \vee J(\ulcorner J\ulcorner\bot\urcorner\urcorner)$. We leave the verification of these results to the reader,[28] but we find this differential role of the principle of coherence in the impossibility results interesting *per se* and worth pursuing in future work. Although not all principles lead against the context of the base theory without CO to straightforward contradictions, or even to 'a' Catrin being justified in a contradiction, we still find the consequences of all of these principles problematic. If 'a' Catrin notices that, in the context of B, taking **GH** as an additional axiom would allow her to prove $J\ulcorner\bot\urcorner$, then she rightly concludes that **GH** should not be taken as an axiom.

7 In closing

In this paper we have investigated the concept of pure justified belief. We tried to lay down basic laws of the logic of justified belief. This resulted in the base theory B. We discussed several additional principles for this logic and provided an overview of inconsistency and consistency results concerning justified belief, in the spirit of what Friedman and Sheard (1987) did for truth. Moreover, we showed that most of the additional principles that have been considered in the literature cannot consistently be added to B, and that the only propositional modal logical principle that can be added (**CPI**), is questionable. These results mirror results from Cieśliński, Horsten and Leitgeb about rational subjective probability that is though more fine grained still to some extent comparable to justified belief (Cieśliński et al., 2022). The results in that article therefore invite similar conclusions about the notion of rational subjective probability.

It is tempting to conclude from this that all the additional principles that were discussed in Sect. 4.2 should be either restricted somehow, or rejected outright. Indeed, our findings are largely in harmony with (but independent from) Williamson's anti-luminosity arguments.

One can try to resist this conclusion in several ways:

1. Exclude self-referential sentences from the scope of the principles;

[28] The verification of these results is obtained by a closer inspection of the proofs of the impossibility results in Sect. 4.3.

2. Argue against the base theory.

But neither of these reactions is appropriate.

Self-referential statements such as the knower sentence can be excluded in two ways. One can treat justified belief as an operator instead of as a predicate, or one can impose a type restriction (of the form "where J does not occur in ϕ") on the principles of the logic of justified belief. Both of these strategies result in a limitation of the generality of the principles of the logic of justified belief, and thus conflicts with the universal ambitions of epistemology. Concerning the base system, we have seen how one may be worried that B is beset by a variant of the problem of logical omniscience due to Nec and CL. We have argued that these worries are unfounded. It is true that no non-idealised agent can serve as a model of all theorems of B. But for every finite fragment of B, there is a non-idealised rational agent that models it. In particular, for each of the inconsistency arguments in this article, there are flesh and blood agents who have gone through them step by step. Also Williamson (2000, pp. 116–118) argues that the closure principle is intuitive and should not be rejected, especially, when one formulates it only over the *pertinent* propositions: "If we reject it, in what circumstances can we gain knowledge by deduction?"

A third way to avoid the conclusion that all the principles that were considered in Sect. 4.2 are unacceptable, consists in restricting the background logic. For instance, we know from the literature on truth theories that no go-theorems tend to melt away when the background logic is weakened from classical logic to partial logic (Halbach & Horsten, 2006). An attempt to partial logic to epistemological notions can, for example, be found in (Schuster, 2022) or (Horsten, 1998).

Another non-classical approach can be found in (Zardini, 2020). Whether some such move might be an acceptable route for epistemologists, is a moot question. One possible ground for scepticism on this score is that in virtually all the literature in contemporary epistemology, full classical logic is assumed.

Acknowledgements We are grateful to two anonymous reviewers of Synthese for their helpful comments and suggestions. Special thanks to Cezary Cieśliński, Jan Heylen and the members of the doctoral colloquium organized by Leon Horsten and Carolin Antos for the fruitful discussion and their helpful suggestions in the colloquium and afterwards.

Funding Open Access funding enabled and organized by Projekt DEAL. Daniela Schuster thanks the Studienstiftung des deutschen Volkes and the Baden-Württemberg Stiftung for financial support.

Declarations

Conflict of interest The authors have no conflict of interest.

References

Anderson, C. A. (1983). The paradox of the knower. *The Journal of Philosophy, 80*(6), 338–355. https://doi.org/10.2307/2026335

Artemov, S. (2008). The logic of justification. *Review of Symbolic Logic, 1*, 477–513.

Boolos, G. (1995). Introductory note to *1951. In Feferman, S., et al. (eds) *Kurt Gödel. Collected works* (Vol. III). Oxford University Press.

Burge, T. (1978). Buridan and epistemic paradox. *Philosophical Studies, 34*(1), 21–35. https://doi.org/10.1007/BF00364686

Burge, T. (1997). Interlocution, perception, and memory. *Philosophical Studies, 86*, 21–47.

Cieśliński, C. (2017). *Deflationism and its logic: The epistemic lightness of truth*. Cambridge University Press.

Cieśliński, C. (unpublished). Believability theories. corrigendum to: The epistemic lightness of truth. Deflationism and its logic. http://cieslinski.filozofia.uw.edu.pl/Corrigendum.pdf

Cieśliński, C., Horsten, L., Leitgeb, H. (2022). Axioms for typefree subjective probability. arXiv:2203.04879

Cross, C. B. (2001). A theorem concerning syntactical treatments of nonidealized belief. *Synthese, 129*(3), 335–341.

Égré, P. (2005). The knower paradox in the light of provability interpretations of modal logic. *Journal of Logic, Language and Information, 14*(1), 13–48.

Feferman, S., Solovay, R. (1990). Note to 1972a. In Feferman, S., et al. (eds) *Kurt Gödel. Collected works* (Vol. II). Oxford University Press.

Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic, 33*, 1–21.

Halbach, V. (1996). *Axiomatische Wahrheitstheorien*. Logica Nova, De Gruyter. https://doi.org/10.1515/9783050072258; http://gbv.eblib.com/patron/FullRecord.aspx?p=4008493

Halbach, V., & Horsten, L. (2006). Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic, 71*, 677–712.

Heylen, J. (2016). Being in a position to know and closure. *Thought: A Journal of Philosophy, 5*(1), 63–67.

Heylen, J. (2020). Rosenkranz's logic of justification and unprovability. *Journal of Philosophical Logic, 49*, 1243–1256.

Hintikka, J. (1962). *Knowledge and belief*. Cornell University Press.

Horsten, L. (1998). A Kripkean approach to unknowability and truth. *Notre Dame Journal of Formal Logic, 39*(3), 389–405. https://doi.org/10.1305/ndjfl/1039182253

Horsten, L. (2011). *The Tarskian turn: Deflationism and axiomatic truth*. MIT Press.

Kant, I. (1993). *Critique of pure reason* (Trans. by Smith, N. K., Second Edn.). Macmillan.

Kaplan, D., & Montague, R. (1960). A paradox regained. *Notre Dame Journal of Formal Logic, 1*(3), 79–90. https://doi.org/10.1305/ndjfl/1093956549

Koons, R. C. (1992). *Paradoxes of belief and strategic rationality*. Cambridge University Press.

Kuznets, R. (2008). *Complexity issues in justification logic*. City University of New York.

Łełyk, M. (unpublished). Universal properties of truth. Research Note.

Lenzen, W. (1980). *Glauben*. Springer.

Myhill, J. (1960). Some remarks on the notion of proof. *The Journal of Philosophy, 57*(14), 461–471.

Rosenkranz, S. (2016). Being in a position to know and closure: Reply to Heylen. *Thought: A Journal of Philosophy, 5*, 68–72.

Rosenkranz, S. (2018). The structure of justification. *Mind, 127*(506), 309–338.

Rosenkranz, S. (2021). *Justification as ignorance: An essay in epistemology*. Oxford University Press.

Schuster, D. (2022). The fixed points of belief and knowledge. Manuscript, under review.

Smith, M. (2018). The logic of epistemic justification. *Synthese, 195*, 3857–3875.

Smith, M. (2022). Is ¬K¬KP a luminous condition? *Asian Journal of Philosophy, 1*(1), 1–10.

Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research, 85*(2), 273–300.

Thomason, R. H. (1980). A note on syntactical treatments of modality. *Synthese, 44*(3), 391–395. https://doi.org/10.1007/BF00413468

Van Fraassen, B. C. (2019). Study of a self-transparent believer. Retrieved October 12, 2019 from https://basvanfraassensblog.home.blog

Van Fraassen, B. C. (2022). Logic of a self-transparent believer. *Filosofiska Notiser, 9*(1), 11–25.

Wang, H. (1974). *From mathematics to philosophy*. Routledge.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford University Press.
Zardini, E. (2020). Closed without boundaries. *Synthese*, 1–39.