Theory Comput Syst (2010) 46: 636–661
DOI 10.1007/s00224-009-9209-3

Self-Referential Justifications in Epistemic Logic

Roman Kuznets

Published online: 7 April 2009

Abstract This paper is devoted to the study of self-referential proofs and/or justifications, i.e., valid proofs that prove statements about these same proofs. The goal is to investigate whether such self-referential justifications are present in the reasoning described by standard modal epistemic logics such as S4. We argue that the modal language by itself is too coarse to capture this concept of self-referentiality and that the language of justification logic can serve as an adequate refinement. We consider well-known modal logics of knowledge/belief and show, using explicit justifications, that S4, D4, K4, and T with their respective justification counterparts LP, JD4, J4, and JT describe knowledge that is self-referential in some strong sense. We also demonstrate that self-referentiality can be avoided for K and D.

In order to prove the former result, we develop a machinery of minimal evidence functions used to effectively build models for justification logics. We observe that the calculus used to construct the minimal functions axiomatizes the reflected fragments of justification logics. We also discuss difficulties that result from an introduction of negative introspection.

Keywords Self-referentiality · Justification logic · Epistemic modal logic · Logic of Proofs

1 Introduction

The concept of self-reference, or self-referentiality, is a recurring topic in epistemology and beyond, with Cantor's Diagonalization Method, Russell's Paradox, and Gödel's Incompleteness Theorems being only a few examples where the phenom-

Supported by Swiss National Science Foundation grant 200021-117699.

R. Kuznets (✉)
Institut für Informatik und angewandte Mathematik, Universität Bern, Neubrückstrasse 10, 3012 Bern, Switzerland
e-mail: kuznets@iam.unibe.ch

enon manifests itself as an object under and/or a tool of investigation. Often self-referentiality is used to demonstrate contradictions or paradoxes, for which reason it is regarded with suspicion and conscious efforts are made to avoid it.

In the framework of formal epistemology, the first question to be clarified is what kinds of self-referring objects are being considered. The case of self-referring sentences in the context of Peano Arithmetic received a thorough treatment in [22]. Moreover, it turns out that many statements about self-reference can be formulated in the modal language, where proofs, Gödel numbers, and the like are abstracted away and concealed in the $\Box$ of modal logic GL.

In this paper, we are also interested in self-referentiality in the context of proofs or justifications, but the self-referring objects we study are proofs themselves.

Following Smoryński, we use the term *self-reference* when talking about self-referring statements, whereas the term *self-referentiality* is reserved for proofs and is the main object of study in this paper.

The question we are trying to answer can be broadly formulated as follows:

> Do we normally use proofs that refer to themselves in mathematical discourse? If so, can we eliminate such self-referential proofs: in a way, can we make proofs *predicative*? If in general this is not possible, then which statements can still be derived without the use of self-referentiality?

Naturally, to answer this question we need to narrow it down, leaving the general discussion to philosophers. It should be clear that the answer strongly depends on the context and in certain contexts can be trivial.

For instance, asking this question about proofs in Peano Arithmetic with the standard Gödel numbering leads to an easy negative answer. Indeed, the Gödel number of a proof is always strictly greater than the Gödel numbers of any parts of the proven statement. Therefore, a valid proof can never be present in the statement proven by this proof.

In contrast, arguments understood in a broader sense as valid reasoning templates or schemes suggest that we can apply them to anything including themselves. For instance, a proof of the shortest tautology, $F \to F$, in any formal system can certainly be applied to any F even if it contains this very proof.

In provability logic, it makes sense to ask whether the use of self-referential proofs is a necessary condition of validity for a particular theorem. Similarly, if one considers knowledge to be justified true belief (see, e.g., [10, 12, 13, 18]), it makes sense to ask whether the use of self-referential justifications is necessary for a particular epistemic fact to be valid. This question is hard to formulate within the confines of the modal language although we have some vague intuitive understanding of its relevance.

Example 1 Consider modal statement

$$\Phi = \neg\Box\neg(P \to \Box P), \tag{1}$$

which is valid in epistemic modal logic S4, i.e., valid for a knowledge agent[1] with positive introspection. This formula intuitively says that it is impossible to know that P does not imply the knowledge of P. Why? If the agent knew that P did not imply the knowledge of P, then

(A) the agent would know that P must be true since otherwise false statement P would imply anything, including $\Box P$, and
(B) the agent would know that she does not know P since otherwise true statement $\Box P$ would follow from anything, including P.

Knowledge is supposed to be factive, so the knowledge of her ignorance of P would mean that the agent indeed would not know P. In summary, the agent would know P without knowing it, a contradiction that shows the impossibility of the supposition that the agent knows that P does not imply the knowledge of P, i.e., in formulas:

1. $\Box\neg(P \to \Box P) \to \Box P$;
2. $\Box\neg(P \to \Box P) \to \Box\neg\Box P$;
3. $\Box\neg\Box P \to \neg\Box P$;
4. $\Box\neg(P \to \Box P) \to \neg\Box P$.

From 1. and 4., $\neg\Box\neg(P \to \Box P)$ follows by propositional reasoning. There seems to be a flavor of self-referentiality in this derivation. Indeed, the knowledge of P here is derived from the knowledge of some fact, $\neg(P \to \Box P)$, that involves $\Box P$, the knowledge of P.

Unfortunately, it is not always clear in the modal language whether this knowledge and that knowledge of the same statement are, in fact, related. An exposition of this phenomenon in Kripke's Red Barn Example can be found in [3].

In Example 1, there is a reason to believe that self-referentiality does occur because the consequents of 1. and 4. have to be related in order to derive $\neg\Box\neg(P \to \Box P)$. The consequent of 4. comes from the consequent of 3., which in its turn follows from part of 3.'s antecedent. Finally, this antecedent, according to 2., is directly related to $\Box P$ inside the parenthesis.

We leave it to the reader to decide how persuasive this argument of the presence of self-referentiality is. But even if it is, there is always a counterargument that there may be another derivation that does not exploit self-referentiality.

This example shows that to study the impact of self-referentiality as a proof technique, we need a richer language that would allow for a finer analysis. In this paper, we show that the language of justification logic (see [3]) fits the bill in many cases. Instead of using statements $\Box F$ (*there exists a proof of* F) wherein proofs are concealed, justification logics employ the construct $t:F$ (read *term* t *serves as a justification for or proof of* F) with proofs explicitly present, which greatly simplifies the task of truth-tracking.

In the justification language, it is easy to see when self-referentiality occurs: when a term t proves something about itself, i.e.,

$$\vdash t:F(t). \tag{2}$$

[1] A knowledge agent only knows true facts, unlike a belief agent whose beliefs can be false.

This is the simplest but not the only type of self-referentiality; for instance, it could happen that $\vdash t_1 : F(t_2)$ and $\vdash t_2 : F(t_1)$, with one proof referring to the other and vice versa. We will discuss both the one-step, or *direct*, self-referentiality and the multi-step one.

Before defining justification logics and plunging into technicalities, we have to explain what effect our results about the justification language have on more familiar epistemic modal logics, such as S4. There is a clear connection between the modal language and the language with explicit justifications:

Definition 2 *Forgetful projection* $^\circ$ turns each justification formula into a modal one by replacing each occurrence of a justification term by $\Box$, $(t:G)^\circ = \Box(G^\circ)$, while commuting with Boolean connectives, $(F \to G)^\circ = F^\circ \to G^\circ$, and keeping sentence letters and Boolean constants intact, $P^\circ = P$ and $\bot^\circ = \bot$.[2]

The forgetful projection of a set X of justification formulas is a set of modal formulas $X^\circ = \{F^\circ \mid F \in X\}$.

A logic L can be viewed as the set of L-theorems. Then, a modal logic ML is said to be the *forgetful projection* of a justification logic JL if $\mathsf{JL}^\circ = \mathsf{ML}$.

It was shown in [1] that the forgetful projection of the first justification logic, Logic of Proofs LP, is exactly S4, i.e., $\mathsf{LP}^\circ = \mathsf{S4}$ (see also [2]). This statement is typically called the Realization Theorem and embodies two directions:

1. Replacing each justification term in an LP-theorem by $\Box$ yields an S4-theorem.
2. Vice versa, it is possible to *realize* all occurrences of $\Box$ in an S4-theorem by justification terms in such a way that the resulting justification formula is valid. This process of restoring terms hidden in $\Box$'s is called *realization*.

For each of modal logics K, D, T, K4, D4, S4, K5, K45, KD45, and S5, a justification counterpart has been developed so that its forgetful projection is exactly this modal logic (see [1, 3, 5, 20, 21]). In this respect, each of these justification logics is a fair representation of its forgetful projection. So the role of self-referentiality in a particular type of reasoning represented by a modal logic, say S4, can be investigated through its justification counterpart, in this case LP.

Definition 3 We say that modal reasoning in a modal logic ML, as represented by its justification counterpart JL, is *not directly self-referential* if each modal theorem G of ML can be realized by a justification theorem G^r that can be derived in JL without using any self-referential statements $t : F(t)$.

The reasoning of ML and JL is *not self-referential* if the realization of each modal theorem G can be achieved without using any cycles of references, such as

$$t_2 : F_1(t_1), \quad \ldots, \quad t_n : F_{n-1}(t_{n-1}), \quad t_1 : F_n(t_n). \tag{3}$$

[2] At this point, the structure of justification terms is not important since forgetful projection erases the terms, structure notwithstanding. The structure of terms, which is the main truth-tracking tool, will be discussed in detail in the next section.

In this paper,[3] we consider several representative examples and show that in all the cases

– either direct self-referentiality is required already on the level of atomic justifications (S4/LP, D4/JD4, K4/J4, and T/JT)
– or self-referentiality can be avoided (K/J and D/JD).

Section 2 describes several justification logics and their forgetful projections. Epistemic semantics for the justification logics from Sect. 2, the so-called F-models, is described in Sect. 3. In Sect. 4, we introduce $*$-calculi, an important tool for constructing F-models. Using the $*$-calculi to construct F-countermodels, in Sect. 5 we prove that the Realization Theorem for S4, D4, K4, and T requires direct self-referentiality. Section 6 demonstrates how to avoid self-referentiality while realizing logics K and D. In Sect. 7, we discuss the difficulties presented by negative introspection. Section 8 outlines directions for future research.

2 Justification Logics

The historically first justification logic, LP, was introduced in [1], where its forgetful projection was shown to be S4 (see also [2]). Justification counterparts for K, D, T, K4, and D4 were developed and the Realization Theorem for them was proved in [5]. The realizations of several modal logics with negative introspection were considered in [3, 20, 21]. These logics with negative introspection present substantial difficulties in applying our methods, which is discussed in detail in Sect. 7. In Sects. 2–6, we focus on modal logics

$$\mathsf{K},\quad \mathsf{D},\quad \mathsf{T},\quad \mathsf{K4},\quad \mathsf{D4},\quad \mathsf{S4} \tag{4}$$

and their respective justification counterparts

$$\mathsf{J},\quad \mathsf{JD},\quad \mathsf{JT},\quad \mathsf{J4},\quad \mathsf{JD4},\quad \mathsf{LP}. \tag{5}$$

The language of justification logic is that of propositional logic enriched by a new construct $t:F$, where F is any formula and t is a justification term:

$$\begin{aligned} F &::= P \mid \bot \mid (F \to F) \mid t:F, \\ t &::= c_i \mid x \mid (t \cdot t) \mid (t + t) \mid !t, \end{aligned}$$

where c_i is a constant from a family of justification constants $c_1, c_2, \ldots, c_n, \ldots$; x is a justification variable; and P is a sentence letter. Constants from the same family are denoted by the same letter with different integer indices; different families are denoted by different letters. ! is a unary operation while $+$ and $\cdot$ are binary operations on terms.[4]

All the six justification logics from (5) share the following axioms and rules:

[3] An earlier version of this paper appeared in conference proceedings [17]. Results for S4/LP date back to [4, 15].

[4] Operation ! is used only in J4, JD4, and LP.

Table 1

Modal scheme	Justification scheme	Name of justification scheme	Is added in logics
$\Box F \to F$	$t{:}F \to F$	A4. Factivity	JT, LP
$\Box F \to \Box\Box F$	$t{:}F \to !t{:}(t{:}F)$	A5. Positive Introspection	J4, JD4, LP
$\Box\bot \to \bot$	$t{:}\bot \to \bot$	A7. Consistency	JD, JD4

A1. Classical propositional axioms[5] and rule *modus ponens*.
A2. *Application Axiom* $s{:}(F \to G) \to (t{:}F \to (s \cdot t){:}G)$.
A3. *Monotonicity Axiom* $s{:}F \to (s+t){:}F$, $t{:}F \to (s+t){:}F$.
R4. *Axiom Internalization Rule*: $\dfrac{}{c_n{:}c_{n-1}{:}\ldots{:}c_1{:}A}$, where A is an axiom, $c_1, \ldots, c_n$ is an initial segment of a family of justification constants.

These axioms and rules alone yield the basic justification logic J, whose forgetful projection is K, the weakest normal modal logic. It is easy to see that the forgetful projection of axioms of J yields theorems of K. Just like the other modal logics from (4) are obtained by adding axiom schemes to K, so can their justification counterparts from (5) be obtained by adding corresponding justification schemes to J. In each case, the added modal axiom scheme is the forgetful projection of the respective justification scheme (see Table 1).[6] It is important to note that the modal Seriality Axiom in the last row of the table is a single axiom, whereas its realization requires an axiom scheme A7.

Theorem 4 (Realization Theorem) [1, 5]

$$\mathsf{J}^\circ = \mathsf{K}, \qquad \mathsf{JD}^\circ = \mathsf{D}, \qquad \mathsf{JT}^\circ = \mathsf{T},$$
$$\mathsf{J4}^\circ = \mathsf{K4}, \qquad \mathsf{JD4}^\circ = \mathsf{D4}, \qquad \mathsf{LP}^\circ = \mathsf{S4}.$$

For each justification logic, a family of weaker logics with restricted rule R4 is defined. Note that this rule has a different scope in different justification logics because they have different sets of axioms. Thus, the following definition of a constant specification depends on the respective logic. In particular, a constant specification for LP may not be a constant specification for J.

Definition 5 A *constant specification* CS *for a justification logic* JL is any set of formulas $c_n{:}c_{n-1}{:}\ldots{:}c_1{:}A$ that can be introduced by Axiom Internalization Rule R4 of this logic. The only requirement is for such a set to be *downward closed*, i.e., if $c_n{:}c_{n-1}{:}\ldots{:}c_1{:}A \in \mathsf{CS}$, then $c_{n-1}{:}\ldots{:}c_1{:}A \in \mathsf{CS}$.

[5]It is typically required that this axiomatization be arranged into finitely many axiom schemes, which is necessary for decidability and complexity results. Since this additional requirement plays no role for self-referentiality, we omit it here.

[6]Axiom and rule numbering is mostly inherited from [3].

Definition 6 Let CS be a constant specification for a justification logic JL. By $\mathsf{JL}_{\mathsf{CS}}$ we understand the logic obtained by replacing R4 in logic JL by rule

$\mathrm{R4}_{\mathsf{CS}}$. *Relativized Axiom Internalization Rule* $\dfrac{c_n{:}\ldots{:}c_1{:}A \in \mathsf{CS}}{c_n{:}\ldots{:}c_1{:}A}$.

Each logic JL from (5) is essentially $\mathsf{JL}_{\mathsf{TCS}}$ with the *total constant specification* TCS:

$$\mathsf{TCS} = \left\{ c_n{:}\ldots{:}c_1{:}A \;\middle|\; \begin{array}{l} A \text{ is an axiom,} \quad c_1, \ldots, c_n \text{ is an initial segment} \\ \text{of a family of justification constants} \end{array} \right\}.$$

This will enable us to treat only the case of $\mathsf{JL}_{\mathsf{CS}}$ in future definitions, formulations, and proofs and not to mention the case of JL explicitly since JL is an instance of $\mathsf{JL}_{\mathsf{CS}}$.

Note 7 Justification logics with axiom A5, e.g., J4, JD4, and LP, allow for a simpler formulation of rule R4, and consequently of a constant specification, of rule $\mathrm{R4}_{\mathsf{CS}}$, and of TCS:

R4′. *Axiom Internalization Rule* $\dfrac{}{c_1{:}A}$.

The purpose of rule R4 is to realize $\underbrace{\Box\ldots\Box}_{n} A$ for any $n > 0$ and any axiom A. Operations on justifications take care of extending the realization to all theorems. But axiom A5, together with rule R4′, enables us to use

$$\underbrace{!\ldots!}_{n-1} c_1{:}\ldots{:}!!c_1{:}!c_1{:}c_1{:}A$$

for the same purpose. The two approaches are largely equivalent, where $c_n \leftrightsquigarrow \underbrace{!\ldots!}_{n-1} c_1$ provides a translation between them. Originally, all the six logics from (5) were formulated with R4′ in [1, 2, 5]. The formulation in this paper follows [3].

Definition 8 A constant specification CS for a justification logic is called

- *self-referential* if $\{a_1 : A_1(b_{i_1}), b_1 : A_2(c_{i_2}), \ldots, e_1 : A_n(a_{i_n})\} \subseteq \mathsf{CS}$, where a, b, c, …, e represent families of constants and axioms $A_j(d_{i_j})$ must have at least one occurrence of constant d_{i_j} from family d;
- *directly self-referential* if $c_1 : A(c_i) \in \mathsf{CS}$;
- *axiomatically appropriate*[7] if
  1. every axiom A of the logic has at least one family of constants c such that $c_1 : A \in \mathsf{CS}$; and
  2. CS is upward closed, i.e., if $c_n{:}\ldots{:}c_1{:}A \in \mathsf{CS}$, then $c_{n+1}{:}c_n{:}\ldots{:}c_1{:}A \in \mathsf{CS}$.

[7] The term is due to Melvin Fitting (see [8]).

These definitions of self-referential and directly self-referential CS use the downward closure of constant specifications:

$$a_n{:}\ldots{:}a_1{:}A(b_i) \in \mathsf{CS} \quad \Longrightarrow \quad a_1{:}A(b_i) \in \mathsf{CS},$$

i.e., if family a refers to family b, this reference happens already on the level of a_1, the first constant in family a. So self-referentiality means the existence of a cycle of references between families of constants, whereas direct self-referentiality requires some family of constants to refer to itself. These two types of self-referentiality are atomic-level manifestations of our general definition of (direct) self-referentiality from Definition 3. As it turns out, already this basic level is often necessary.

The following property is fundamental for justification logics and is an important tool in proving the Realization Theorem.

Lemma 9 (Internalization Property) [1] *Let* $\mathsf{JL_{CS}}$ *be a justification logic with an axiomatically appropriate* CS. *Then, for any derivation* $F_1, \ldots, F_n \vdash_{\mathsf{JL_{CS}}} B$ *there exists an evidence term* $s(x_1, \ldots, x_n)$ *such that*

$$t_1{:}F_1, \ldots, t_n{:}F_n \vdash_{\mathsf{JL_{CS}}} s(t_1, \ldots, t_n){:}B. \tag{6}$$

Proof A step-by-step translation from the given derivation into the target one.

A	$\rightsquigarrow$	$c_1{:}A$	($\mathsf{R4_{CS}}$), where A is an axiom, the existence of $c_1{:}A \in \mathsf{CS}$ is guaranteed by axiomatic appropriateness
F_i	$\rightsquigarrow$	$t_i{:}F_i$	(hypothesis)
$c_n{:}\ldots{:}c_1{:}A$	$\rightsquigarrow$	$c_{n+1}{:}c_n{:}\ldots{:}c_1{:}A$	($\mathsf{R4_{CS}}$), where $c_n{:}\ldots{:}c_1{:}A \in \mathsf{CS}$, again using axiomatic appropriateness
$\dfrac{D \to G \quad D}{G}$	$\rightsquigarrow$	$\dfrac{s_1{:}(D \to G) \quad s_2{:}D}{(s_1 \cdot s_2){:}G}$	using A2 and *modus ponens* twice

□

Total constant specification TCS is always directly self-referential. Therefore, the standard proofs of the Realization Theorem from [2, 5] only show that realization is possible when direct self-referentiality is used. Our first task is to determine when realization cannot be achieved without (directly) self-referential CS. A relationship to Definition 3 can be described by the following

Proposition 10 *Let a modal logic* ML *be the forgetful projection of a justification logic* JL, *i.e.*, $\mathsf{JL}^\circ = \mathsf{ML}$.

1. *If* $(\mathsf{JL_{CS}})^\circ \neq \mathsf{ML}$ *for any* CS *that is not directly self-referential*, ML/JL *describe directly self-referential reasoning.*
2. *If* $(\mathsf{JL_{CS}})^\circ \neq \mathsf{ML}$ *for any* CS *that is not self-referential*, ML/JL *describe self-referential reasoning.*

3 Epistemic Models for Justification Logics

The self-referentiality of S4, D4, K4, and T is established by a semantic argument. The Kripke-like models we use, epistemic F-models, were first developed by Fitting

for LP. The proof of soundness and completeness of LP with respect to them, as well as their adaptation to J, JT, and J4 can be found in [8]. Soundness and completeness arguments for J and JD can be found in [20], for JT and J4 in [3]. The F-models for JD4 are, perhaps, first developed here.

Definition 11 (F-models) An *F-model* is a quadruple $\mathcal{M} = \langle W, R, \mathcal{A}, V\rangle$, where $\langle W, R, V\rangle$ is a Kripke model with

- a set of worlds $W \neq \emptyset$,
- an accessibility relation $R \subseteq W \times W$, and
- a valuation function $V : SLet \to 2^W$ that assigns to a sentence letter P a set $V(P) \subseteq W$ of all worlds where this sentence letter is deemed true.

Finally, an *admissible evidence function* $\mathcal{A} : Tm \times Fm \to 2^W$ assigns to a pair of a term t and a formula F a set $\mathcal{A}(t, F) \subseteq W$ of all worlds where t is deemed admissible evidence for F. Depending on the logic, there are various restrictions on the types of R and $\mathcal{A}$ allowed. The following closure conditions must be satisfied by $\mathcal{A}$ for all justification logics:

C2. $\mathcal{A}(t, F \to G) \cap \mathcal{A}(s, F) \subseteq \mathcal{A}(t \cdot s, G)$;
C3. $\mathcal{A}(t, F) \cup \mathcal{A}(s, F) \subseteq \mathcal{A}(t + s, F)$;
CS. $\mathcal{A}(c_n, c_{n-1}{:}\ldots{:}c_1{:}A) = W$, where $n \geq 1$ and $c_n{:}c_{n-1}{:}\ldots{:}c_1{:}A \in \mathsf{CS}$.

The forcing relation $\Vdash$ is defined as follows:

- $\mathcal{M}, w \Vdash P$ iff $w \in V(P)$, where P is a sentence letter;
- Boolean cases are standard;
- $\mathcal{M}, w \Vdash t{:}F$ iff (1) $\mathcal{M}, u \Vdash F$ for all wRu and (2) $w \in \mathcal{A}(t, F)$.

Closure conditions C2 and C3 are required to validate axioms A2 and A3 respectively, which is reflected in their numbering. The additional conditions depend on the axioms added to the logic:

- for JT_CS and LP_CS, axiom $t{:}F \to F$ requires that R be *reflexive*.
- For JD_CS and $\mathsf{JD4}_\mathsf{CS}$, axiom $t{:}\bot \to \bot$ requires that R be *serial*.
- For $\mathsf{J4}_\mathsf{CS}$, $\mathsf{JD4}_\mathsf{CS}$, and LP_CS, axiom $t{:}F \to !t{:}t{:}F$ requires that R be *transitive*. In addition, two more closure conditions are imposed on $\mathcal{A}$:

  C5. $\mathcal{A}(t, F) \subseteq \mathcal{A}(!t, t{:}F)$;
  Monotonicity. wRu and $w \in \mathcal{A}(t, F)$ imply $u \in \mathcal{A}(t, F)$.

Note that $w \in \mathcal{A}(t, F)$ in no way implies that F itself holds at w. Rather, $w \in \mathcal{A}(t, F)$ means that at world w term t is acceptable, although not necessarily conclusive, evidence for F.

Just as we sometimes talk about an accessibility relation on W or a valuation function on W without presenting the whole Kripke model, we will often deal with admissible evidence functions without presenting a specific F-model. Note that due to the Monotonicity Condition, to determine whether something is an admissible evidence function, at least for some justification logics, we need to know both W and R; hence, we will usually talk about admissible evidence functions on a given Kripke frame $\langle W, R\rangle$.

Theorem 12 (Completeness Theorem) [3, 8, 16, 20] *Justification logics* $\mathsf{J_{CS}}$, $\mathsf{JT_{CS}}$, $\mathsf{J4_{CS}}$, *and* $\mathsf{LP_{CS}}$ *are sound and complete with respect to their F-models.* $\mathsf{JD_{CS}}$ *and* $\mathsf{JD4_{CS}}$ *are sound with respect to their F-models*; *completeness also holds provided* CS *is axiomatically appropriate.*

Proof The proof is by the standard maximal consistent set construction. The complete details can be found in [16]. □

The method we employ to prove direct self-referentiality consists of taking a modal theorem, such as (1) or its variant for weaker logics, and showing that it cannot be realized unless directly self-referential constants are used. To show the impossibility, we take a possible realization and construct a countermodel for it under the assumption that constants are not directly self-referential. So at the center of our method is our ability to construct a model with given properties. The main challenge, of course, lies in constructing the admissible evidence function with given properties since creating an underlying Kripke model is a routine procedure from modal logic.

4 ∗-Calculi and Minimal Evidence Functions

The method for constructing the so-called *minimal* admissible evidence functions[8] goes back to [19] but was first explicitly shaped as a calculus, for the case of LP, in [14].

Definition 13 Let $\mathcal{F} = \langle W, R\rangle$ be a Kripke frame. A *possible evidence function on* $\mathcal{F}$ is any function $\mathcal{B}: Tm \times Fm \to 2^W$.

We will use possible evidence functions to formulate positive conditions on the admissible evidence function we plan to construct: namely, to describe which terms have to be evidence for which formulas. Note also that an admissible evidence function on $\mathcal{F}$ is, by definition, also a possible evidence function on $\mathcal{F}$.

Definition 14 For a given Kripke frame $\mathcal{F} = \langle W, R\rangle$, we say that a possible evidence function $\mathcal{B}_2$ on $\mathcal{F}$ is *based* on a possible evidence function $\mathcal{B}_1$, also on $\mathcal{F}$, and write $\mathcal{B}_1 \subseteq \mathcal{B}_2$ if $\mathcal{B}_1(t, F) \subseteq \mathcal{B}_2(t, F)$ for any term t and any formula F.

Intuitively, $\mathcal{B} \subseteq \mathcal{A}$ means that admissible evidence function $\mathcal{A}$ satisfies the positive conditions set forth in $\mathcal{B}$. The goal is typically to construct the minimal admissible evidence function based on the given possible evidence function $\mathcal{B}$:

Definition 15 Let $\mathcal{B}$ be a possible evidence function on a Kripke frame $\mathcal{F} = \langle W, R\rangle$. The *minimal* admissible evidence function $\mathcal{A}$ based on $\mathcal{B}$ must satisfy two conditions:

[8] Perhaps, it would be more accurate to call them the *smallest* admissible evidence functions, but the term *minimal* has been traditionally used and we leave it here for the sake of consistency.

1. it is based on $\mathcal{B}$, i.e., $\mathcal{B} \subseteq \mathcal{A}$;
2. it is the smallest one, i.e., $\mathcal{B} \subseteq \mathcal{A}' \Longrightarrow \mathcal{A} \subseteq \mathcal{A}'$ for any other admissible evidence function $\mathcal{A}'$ on the same Kripke frame.

Here are the axioms and rules of the calculi that describe minimal evidence functions[9]. We collectively call the two types of axiom systems described below—$*_{\mathsf{CS}}$ and $*!_{\mathsf{CS}}$—the $*$-calculi.

Definition 16 ($*$-Calculi) Let CS be a constant specification for one of the justification logics from (5). The axioms and rules of $*_{\mathsf{CS}}$-calculus for logics J_{CS}, $\mathsf{JD}_{\mathsf{CS}}$, and $\mathsf{JT}_{\mathsf{CS}}$ are as follows:

$*\mathsf{CS}$. Axioms: $*(c_n,\ c_{n-1}{:}\ldots{:}c_1{:}A)$, where $n \geq 1$ and $c_n{:}c_{n-1}{:}\ldots{:}c_1{:}A \in \mathsf{CS}$.

$*$A2. *Application Rule* $\dfrac{*(s,\ F \to G) \qquad *\,(t,\ F)}{*(s \cdot t,\ G)}$.

$*$A3. *Sum Rule* $\dfrac{*(s,\ F)}{*(s+t,\ F)}$, $\dfrac{*(t,\ F)}{*(s+t,\ F)}$.

For the logics with positive introspection, $\mathsf{J4}_{\mathsf{CS}}$, $\mathsf{JD4}_{\mathsf{CS}}$, and $\mathsf{LP}_{\mathsf{CS}}$, an additional rule has to be added:

$*$A5. *Positive Introspection Rule* $\dfrac{*(t,\ F)}{*(!\,t,\ t{:}F)}$.

The resulting calculus is called $*!_{\mathsf{CS}}$-calculus.

The Internalization Property can be reformulated for the $*$-calculi:

Lemma 17 (Internalization Property) *Let* $\mathsf{JL}_{\mathsf{CS}}$ *be a justification logic from* (5) *with an axiomatically appropriate* CS. *Then, for any derivation* $F_1, \ldots, F_n \vdash_{\mathsf{JL}_{\mathsf{CS}}} B$ *and for the evidence term* $s(x_1, \ldots, x_n)$ *constructed for this derivation in Lemma* 9,

$$*(t_1, F_1), \ldots, *(t_n, F_n) \vdash_{*_{\mathsf{CS}}} *(s(t_1, \ldots, t_n), B). \tag{7}$$

Proof The proof repeats that of Lemma 9, only ($\mathsf{R4}_{\mathsf{CS}}$) in the target derivation should be replaced by $*\mathsf{CS}$, while A2 followed by double *modus ponens* becomes $*$A2. □

Note that neither $*$A3 nor $*$A5 is used in this proof. This is the reason we can use $*_{\mathsf{CS}}$ even for logics with positive introspection.

The converse statement does not hold for $*!_{\mathsf{CS}}$ as the following example[10] shows:

Example 18 $*(x, P) \vdash_{*!_{\mathsf{CS}}} *(!\,x, x{:}P)$ for any CS, but surely $P \nvdash_{\mathsf{JL}_{\mathsf{CS}}} x{:}P$.

[9]For brevity, we will sometimes omit the word *admissible* and call them *minimal evidence functions*.

[10]The example is due to Vladimir Krupski.

But we can prove a weaker statement:

Lemma 19 *For a justification logic* $\mathsf{JL_{CS}}$ *with positive introspection, i.e., for* $\mathsf{J4_{CS}}$, $\mathsf{JD4_{CS}}$, *and* $\mathsf{LP_{CS}}$, *if*

$$*(t_1, F_1), \ldots, *(t_n, F_n) \vdash_{*!_{\mathsf{CS}}} *(s, B),$$

then

1. $t_1 : F_1, \ldots, t_n : F_n,\ F_1, \ldots, F_n \vdash_{\mathsf{JL_{CS}}} s : B$,
2. $t_1 : F_1, \ldots, t_n : F_n,\ F_1, \ldots, F_n \vdash_{\mathsf{JL_{CS}}} B$.

Proof The proof of both claims is by simultaneous induction on the given $*!_{\mathsf{CS}}$-derivation.

For $*(c_n, F)$, an instance of $*\mathsf{CS}$, it is clear that $c_n : F \in \mathsf{CS}$, where F is either in CS (for $n > 1$) or an axiom (for $n = 1$). Thus, both $c_n : F$ and F are derivable in $\mathsf{JL_{CS}}$.

For hypothesis $*(t_i, F_i)$, both F_i (Claim 2) and $t_i : F_i$ (Claim 1) are taken as hypotheses in our $\mathsf{JL_{CS}}$-derivations.

If $*(s_1 \cdot s_2, G)$ is obtained by $*$A2 from $*(s_1, F \to G)$ and $*(s_2, F)$, then in $\mathsf{JL_{CS}}$ we can derive

1. $(s_1 \cdot s_2) : G$ from $s_1 : (F \to G)$ and $s_2 : F$ and
2. G from $F \to G$ and F.

The case of $*$A3 is similar to $*$A2.

Let $*(!s_1, s_1 : F)$ be obtained from $*(s_1, F)$ by $*$A5. Claim 1 for s_1, which holds by IH, happens to coincide with Claim 2 for $!s_1$: they both require that $s_1 : F$ be derivable. Then Claim 1 for $!s_1$, i.e., derivability of $!s_1 : s_1 : F$, can be inferred from Claim 2 for $!s_1$ by means of positive introspection A5. □

Since in this proof, hypotheses $t_i : F_i$ are needed only for Claim 1, whose proof interacts with the proof of Claim 2 only in the $*$A5-clause, and since $*$A5 is also the only clause where positive introspection is used, it follows that for $*_{\mathsf{CS}}$ the converse of Lemma 17 holds:

Lemma 20 *For a justification logic* $\mathsf{JL_{CS}}$ *without positive introspection, i.e., for* $\mathsf{J_{CS}}$, $\mathsf{JD_{CS}}$, *and* $\mathsf{JT_{CS}}$, *if*

$$*(t_1, F_1), \ldots, *(t_n, F_n) \vdash_{*_{\mathsf{CS}}} *(s, B),$$

then

$$F_1, \ldots, F_n \vdash_{\mathsf{JL_{CS}}} B.$$

Corollary 21 *For any* $\mathsf{JL_{CS}} \in \{\mathsf{J_{CS}}, \mathsf{JD_{CS}}, \mathsf{JT_{CS}}, \mathsf{J4_{CS}}, \mathsf{JD4_{CS}}, \mathsf{LP_{CS}}\}$ *and its corresponding* $*$*-calculus* $\vdash_*$,

$$\vdash_* *(s, B) \quad \Longrightarrow \quad \mathsf{JL_{CS}} \vdash B.$$

In order to define minimal evidence functions in terms of the $*$-calculi, we use the following piece of notation:

Definition 22 For a possible evidence function $\mathcal{B}$ on a Kripke frame $\mathcal{F} = \langle W, R\rangle$ and a world $w \in W$,

$$\mathcal{B}^*_w = \{*(t, F) \mid w \in \mathcal{B}(t, F)\}. \tag{8}$$

So $\mathcal{B}^*_w$ contains $*(t, F)$ iff $w \in \mathcal{B}(t, F)$. In this sense $*$ can be seen as an abbreviation for $w \in \mathcal{B}$.

Theorem 23 *Let $\mathcal{B}$ be a possible evidence function on a Kripke frame $\mathcal{F} = \langle W, R\rangle$. Define possible evidence function $\mathcal{A}$ as follows: for logics* J_{CS}, $\mathsf{JD}_{\mathsf{CS}}$, *and* $\mathsf{JT}_{\mathsf{CS}}$, *let*

$$*(t, F) \in \mathcal{A}^*_w \iff \mathcal{B}^*_w \vdash_{*\mathsf{CS}} *(t, F); \tag{9}$$

for logics $\mathsf{J4}_{\mathsf{CS}}$, $\mathsf{JD4}_{\mathsf{CS}}$, *and* $\mathsf{LP}_{\mathsf{CS}}$, *we assume, in addition, that R is transitive and let*

$$*(t, F) \in \mathcal{A}^*_w \iff \mathcal{B}^*_w \cup \bigcup_{uRw} \mathcal{B}^*_u \vdash_{*!\mathsf{CS}} *(t, F). \tag{10}$$

For each of the six logics, $\mathcal{A}$ so defined is the minimal evidence function based on $\mathcal{B}$.

Proof The $*$-calculi act locally, within each world, as do most closure conditions with the exception of Monotonicity. Since $\mathcal{B}^*_w$ is part of the set of hypotheses in both (9) and (10), clearly $\mathcal{B} \subseteq \mathcal{A}$.

In both cases, $\mathcal{A}$ at each world is built from $\mathcal{B}$ at the same world by applying the closure rules (in the equivalent form of $*$-calculus rules), which have to be satisfied anyway. The additional hypotheses from $\mathcal{B}^*_u$ in (10), where uRw, must be satisfied at w due to the Monotonicity Condition: if $u \in \mathcal{B}(t, F)$, then $u \in \mathcal{E}(t, F)$ for any admissible evidence function $\mathcal{E}$ based on $\mathcal{B}$. It follows by the Monotonicity Condition for $\mathcal{E}$ that $w \in \mathcal{E}(t, F)$. So these additional hypotheses do not violate the minimality of $\mathcal{A}$.

It remains to show that $\mathcal{A}$ is, in fact, an admissible evidence function. Rules $*$A2, $*$A3, and $*$A5 guarantee that closure conditions C2, C3, and C5 respectively are satisfied (note that $*$A5 is only used in (10), where C5 has to be satisfied). The axioms from $*\mathsf{CS}$ similarly take care of the CS Closure Condition. The Monotonicity Condition needs to be satisfied only when (10) is used. Let us show that $w \in \mathcal{A}(t, F)$ whenever $u \in \mathcal{A}(t, F)$ and uRw. $u \in \mathcal{A}(t, F)$ means that

$$\mathcal{B}^*_u \cup \bigcup_{zRu} \mathcal{B}^*_z \vdash_{*!\mathsf{CS}} *(t, F)$$

by definition of $\mathcal{A}$. For these logics we assume R to be transitive, so zRu implies zRw, given uRw. Therefore,

$$\mathcal{B}^*_u \cup \bigcup_{zRu} \mathcal{B}^*_z \subseteq \mathcal{B}^*_w \cup \bigcup_{zRw} \mathcal{B}^*_z$$

so that

$$\mathcal{B}^*_w \cup \bigcup_{zRw} \mathcal{B}^*_z \vdash_{*!_{\mathsf{CS}}} *(t, F),$$

i.e., $w \in \mathcal{A}(t, F)$. □

It should be noted that, apart from being a method for constructing models, the $*$-calculi axiomatize the so-called *reflected fragments* of the respective justification logics.

Definition 24 The *reflected fragment* $\mathsf{rJL}_{\mathsf{CS}}$ of a justification logic $\mathsf{JL}_{\mathsf{CS}}$ consists of all its theorems of form $t:F$:

$$\mathsf{rJL}_{\mathsf{CS}} = \{t:F \mid \mathsf{JL}_{\mathsf{CS}} \vdash t:F\}.$$

In fact, Nikolai Krupski in [14] introduced the $*!$-calculus[11] to axiomatize rLP, the reflected fragment of LP. His result can be extended to other logics as follows:

Theorem 25 [14, 16]

1. *The reflected fragment* $\mathsf{rJL}_{\mathsf{CS}}$ *of* $\mathsf{JL}_{\mathsf{CS}} \in \{\mathsf{J}_{\mathsf{CS}}, \mathsf{JD}_{\mathsf{CS}}, \mathsf{JT}_{\mathsf{CS}}\}$ *is completely axiomatized by the* $*_{\mathsf{CS}}$*-calculus*:

$$\mathsf{rJL}_{\mathsf{CS}} \vdash t:F \iff \mathsf{JL}_{\mathsf{CS}} \vdash t:F \iff *_{\mathsf{CS}}\text{-}calculus \vdash *(t, F).$$

2. *The reflected fragment* $\mathsf{rJL}_{\mathsf{CS}}$ *of* $\mathsf{JL}_{\mathsf{CS}} \in \{\mathsf{J4}_{\mathsf{CS}}, \mathsf{JD4}_{\mathsf{CS}}, \mathsf{LP}_{\mathsf{CS}}\}$ *is completely axiomatized by the* $*!_{\mathsf{CS}}$*-calculus*:

$$\mathsf{rJL}_{\mathsf{CS}} \vdash t:F \iff \mathsf{JL}_{\mathsf{CS}} \vdash t:F \iff *!_{\mathsf{CS}}\text{-}calculus \vdash *(t, F).$$

Proof (Sketch) In each case the middle statement is equivalent to the left one by definition.

To derive the right statement from the left one, we use proof by contradiction. Indeed, if $\nvdash *(t, F)$, then by Theorem 23 it would be possible to construct a model with $w \notin \mathcal{A}(t, F)$ for some world w by taking $\mathcal{A}$ to be the minimal evidence function based on the empty possible evidence function, $\mathcal{B}(t, F) \equiv \emptyset$. This world would then falsify theorem $t:F$ in violation of soundness.

Predictably, to get the left statement from the right one, completeness can be used. If $\vdash *(t, F)$, then $\mathsf{JL}_{\mathsf{CS}} \vdash F$ by Corollary 21. Thus, F is valid by soundness. By Theorem 23, it follows from $\vdash *(t, F)$ that $\mathcal{A}(t, F) = W$ for any admissible evidence function $\mathcal{A}$ in any model. Therefore, $t:F$ is valid and hence derivable by completeness.

Full details of the proof can be found in [16]. □

Armed with minimal evidence functions as a tool for constructing F-models, we are now ready to prove direct self-referentiality.

[11] Under the name of C(CS).

5 Self-Referential Cases: S4, D4, T, and K4

Theorem 26 *Realization of* S4 *in* LP, *of* D4 *in* JD4, *and of* T *in* JT *requires directly self-referential constants and, hence, direct self-referentiality.*

Proof Formula $\Phi = \neg\Box\neg(P \to \Box P)$ from (1) is derivable in all the three modal logics: D4, T, and S4.[12] Indeed, the S4-derivation from Example 1 (on p. 638) uses only normal modal reasoning and the reflection axiom (in line 3). Hence, it can be performed in T as is. Here is a similar derivation for D4:

1. $\Box\neg(P \to \Box P) \to \Box P$ (as in Example 1);
2. $\Box\neg(P \to \Box P) \to \Box\neg\Box P$ (as in Example 1);
3. $\Box\neg(P \to \Box P) \to (\neg\Box P \to \bot)$ (from 1. by propositional reasoning);
4. $\Box\Box\neg(P \to \Box P) \to (\Box\neg\Box P \to \Box\bot)$ (from 3. by normal modal reasoning);
5. $\Box\neg(P \to \Box P) \to (\Box\neg\Box P \to \Box\bot)$ (from 4. by transitivity);
6. $\Box\neg(P \to \Box P) \to (\Box\neg\Box P \to \bot)$ (from 5. by seriality);
7. $\Box\neg(P \to \Box P) \to \bot$ (from 6. and 2. by propositional reasoning).

The last formula is nothing but Φ.

Our goal is to show that no potential realization of Φ can be valid in F-models of $\mathsf{JD4_{CS}}$, $\mathsf{JT_{CS}}$, or $\mathsf{LP_{CS}}$ respectively unless CS contains directly self-referential constants.

Let $\mathsf{JL} \in \{\mathsf{JD4}, \mathsf{JT}, \mathsf{LP}\}$ and CS be the maximal constant specification for JL without directly self-referential constants:

$$\mathsf{CS} = \left\{ c_n : c_{n-1} : \ldots : c_1 : A \;\middle|\; \begin{array}{l} A \text{ is an axiom of JL that does not} \\ \text{contain constants } c_i \text{ from family } c \end{array} \right\}. \tag{11}$$

For any pair of terms t and t' proposed as realizations of the two $\Box$'s in Φ, we construct an F-model for $\mathsf{JL_{CS}}$ that falsifies $\neg t : [\neg(P \to t' : P)]$, thus demonstrating that no realization of Φ is $\mathsf{JL_{CS}}$-valid. Note that only the soundness of $\mathsf{JL_{CS}}$ with respect to its F-models is used in this argument. The additional condition for CS to be axiomatically appropriate, necessary for completeness in case of JD4, thus plays no role, even though it is, in fact, satisfied for the CS from (11).

Given t and t', consider the following F-model for $\mathsf{JL_{CS}}$: $\mathcal{M} = \langle W, R, \mathcal{A}, V\rangle$ with the Kripke frame $\langle W, R\rangle$ that consists of a single reflexive world w, i.e., with $W = \{w\}$ and $R = \{\langle w, w\rangle\}$. Such R is obviously serial, reflexive, and transitive, thus making the frame suitable for JD4, JT, and LP alike. Since w is the only world in the model, we can write

$$\begin{array}{rlll} \Vdash F & \text{instead of} & \mathcal{M}, w \Vdash F, & \\ \mathcal{A}(s, F) & \text{instead of} & w \in \mathcal{A}(s, F), & \text{and} \\ \neg\mathcal{A}(s, F) & \text{instead of} & w \notin \mathcal{A}(s, F). & \end{array}$$

[12] The idea to use this formula for S4 was suggested by an anonymous referee of an earlier version of this paper. Melvin Fitting then conjectured that it could also be used for the other two modal logics.

Let us analyze what is needed to falsify $\neg t{:}[\neg(P \to t'{:}P)]$ (at the only world in the model). Clearly, it is sufficient to satisfy $t{:}[\neg(P \to t'{:}P)]$. So the first requirement on the model is that

$$\mathcal{A}(t,\ \neg(P \to t'{:}P)). \tag{12}$$

In addition, $\neg(P \to t'{:}P)$ itself has to be true. This amounts to two requirements:

$$\Vdash P \tag{13}$$

and $\nVdash t'{:}P$. In general, there are two ways to guarantee the latter: either by making P false in one of the accessible worlds or by making t' not admissible as evidence for P. In our case, the only accessible world is w itself, so (13) effectively prohibits the first path. Thus, we must require

$$\neg\mathcal{A}(t',\ P). \tag{14}$$

Satisfying (12)–(14) is clearly sufficient for our purposes.

Let $V(P) = W = \{w\}$, which takes care of (13). The truth values of the other sentence letters are unimportant. Let $\mathcal{B}$ be a possible evidence function on $\langle W, R\rangle$ defined by

$$\mathcal{B}^*_w = \{*(t,\ \neg(P \to t'{:}P))\}, \tag{15}$$

and let $\mathcal{A}$ be the minimal evidence function based on this $\mathcal{B}$. (Note that $\mathcal{A}$ depends on terms t and t'.) This choice of $\mathcal{A}$ guarantees that (12) is satisfied, so it remains to verify (14), i.e., according to (9) or (10), to show that

$$*(t,\ \neg(P \to t'{:}P)) \nvdash_* *(t',\ P) \tag{16}$$

in the corresponding $*$-calculus. This is achieved by means of the following lemma:

Lemma 27 *For any subterm s of term t':*

1. *If $\vdash_* *(s, F)$, then F does not contain occurrences of t'.*
2. *If $*(t,\ \neg(P \to t'{:}P)) \vdash_* *(s, F)$, but $\nvdash_* *(s, F)$, then F has at least one occurrence of t'. Moreover, if F is an implication, then $F = \neg(P \to t'{:}P)$.*[13]

Here $\vdash_*$ represents $\vdash_{*\mathsf{CS}}$ in the case of JT_CS or $\vdash_{*!\mathsf{CS}}$ in the case of $\mathsf{JD4}_\mathsf{CS}$ and LP_CS.

The proof of this lemma is rather technical and sheds little light on what is going on. Let us first finish the proof of the theorem. The proof of the lemma can be found below on p. 652.

Consider $*(t', P)$. $\mathsf{JL}_\mathsf{CS} \nvdash P$, so by Corollary 21, $\nvdash_* *(s, P)$. Further, since t' does not occur in P, by Lemma 27.2, $*(t,\ \neg(P \to t'{:}P)) \nvdash_* *(s, P)$ either.

Thus, the constructed model satisfies (12)–(14) and, hence, falsifies the proposed realization of (1). □

[13] We consider $\neg G$ to be an abbreviation of $G \to \bot$. Assuming that $\neg$ is a primary connective would only simplify matters.

The technicalities in the formulation of Lemma 27 may obscure the fact that it is nothing but a formal reformulation of the argument in Example 1. In fact, the example was originally inspired by this lemma.

Lemma 17 helps to understand how (16) can be violated: if term t' encodes a derivation $\neg(P \to t' : P) \vdash_{\mathsf{JL_{CS}}} P$. In this derivation, the hypothesis has to be used because P is not valid on its own; this argument corresponds to Corollary 21. And any meaningful way of using hypothesis $\neg(P \to t' : P)$ requires that it be part of an axiom, which is represented in t' by a constant. This constant would justify the axiom and, thus, would refer to t', at the same time being part of t'. This is the essence of the second claim of Lemma 27. The difference between the claims of Lemma 27 is the difference between using a hypothesis "in a meaningful way" (Claim 2) and otherwise (Claim 1).

Proof of Lemma 27

(A) **Case $\mathbf{s = x}$**, a justification variable:
$*(x, F)$ can only be derived from $*(t, \neg(P \to t' : P))$ and only if they coincide; therefore, $t = x$ and $F = \neg(P \to t' : P)$, which does contain t' and is the only allowed implication.

(B) **Case $\mathbf{s = c_n}$**, a justification constant:
Unless $*(c_n, F)$ coincides with the hypothesis as in 5, it can only be derived by $*\mathsf{CS}$, in which case $c_n : F \in \mathsf{CS}$ and we are in the situation of Claim 1. Then, either $n = 1$ and $F = A$ or $n > 1$ and $F = c_{n-1} : \ldots : c_1 : A$, where A is an axiom. Since CS is not directly self-referential, A cannot contain occurrences of c_n, a subterm of t', and neither can $c_1, \ldots, c_{n-1}$. Thus, F does not contain t'.

(C) **Case $\mathbf{s = !\, s_1}$** (only for logics $\mathsf{JD4_{CS}}$ and $\mathsf{LP_{CS}}$):
Unless $*(!\, s_1, F)$ coincides with the hypothesis as in 5, it can only be derived by $*$A5 from $*(s_1, G)$ and only if $F = s_1 : G$. If $*(s_1, G)$ is derivable without the hypothesis (Claim 1), G does not contain t' by IH, whereas s_1 is a proper subterm of t'. Therefore, $F = s_1 : G$ does not contain t'.
If $*(s_1, G)$ can only be derived from the hypothesis (Claim 2), G contains t' by IH, and so does $F = s_1 : G$, which is not an implication.

(D) **Case $\mathbf{s = s_1 + s_2}$**:
Unless $*(s_1 + s_2, F)$ coincides with the hypothesis as in 5, it can only be derived by $*$A3 from $*(s_i, F)$ for some $i = 1, 2$. Therefore, either claim for F holds by IH.

(E) **Case $\mathbf{s = s_1 \cdot s_2}$**:
Unless $*(s_1 \cdot s_2, F)$ coincides with the hypothesis as in 5, it can only be derived by $*$A2 from $*(s_1, G \to F)$ and $*(s_2, G)$ for some formula G. If both premises can be derived without the hypothesis (Claim 1), then $G \to F$ does not contain t' by IH, and consequently neither does F.
It turns out that this is the only possibility. Indeed, if $*(s_2, G)$ is not derivable without the hypothesis, G must contain t' by IH. Therefore, $G \to F$ also contains t', and by IH $*(s_1, G \to F)$ is not derivable without the hypothesis either. Thus, whenever the hypothesis is needed at all, $*(s_1, G \to F)$ definitely requires it. Suppose it does. Being an implication, by IH

$$G \to F = \neg(P \to t' : P) = (P \to t' : P) \to \bot$$

must be the only implication allowed in Claim 2. So $G = P \to t' : P$. Then $*(s_2, G)$ can only belong to Claim 2 because G contains t'. However, G is an implication other than the only one allowed by IH. This contradiction completes the proof of 5. □

Theorem 28 *Realization of* K4 *in* J4 *requires direct self-referentiality.*

Proof Formula Φ from (1) is not derivable in K4 and thus cannot be used here. But since the Hilbert-style axiom system for D4 is obtained from that of K4 by adding just one axiom, Seriality, $\mathsf{K4} \vdash \neg\Box\bot \to \neg\Box\neg(P \to \Box P)$.[14] We will show that its equivalent form

$$\Psi = \Box\neg(P \to \Box P) \to \Box\bot \tag{17}$$

cannot be realized in J4 without directly self-referential constants.

For any potential realization

$$\Psi^r = t : [\neg(P \to t' : P)] \to k : \bot, \tag{18}$$

we construct an F-model for $\mathsf{J4_{CS}}$ that falsifies Ψ^r, thus showing that no realization of Ψ is $\mathsf{J4_{CS}}$-valid. As in Theorem 26, here CS is the maximal constant specification without directly self-referential constants defined by (11) with $\mathsf{JL} = \mathsf{J4}$.

This time the frame in the falsifying model consists of a single irreflexive world, i.e., $W = \{w\}$, $R = \emptyset$. In such a model, any F is vacuously true at all accessible worlds. Therefore, $\Vdash s : F$ iff $\mathcal{A}(s, F)$. Once again, we take $\mathcal{A}$ to be the minimal evidence function based on $\mathcal{B}$ defined by (15). (Note that R is not present in the definition of $\mathcal{B}$, so the fact that R used in Theorem 26 differs from the one used here plays no role as long as W is the same.) Valuation V is not important.

Clearly, $\mathcal{A}(t, \neg(P \to t' : P))$, so to falsify (18) it is sufficient to show $\neg\mathcal{A}(k, \bot)$, which, according to (10), is equivalent to

$$*(t, \neg(P \to t' : P)) \nvdash_{*!_{\mathsf{CS}}} *(k, \bot).$$

Suppose towards a contradiction that $*(t, \neg(P \to t' : P)) \vdash_{*!_{\mathsf{CS}}} *(k, \bot)$. By Lemma 19.2,

$$\neg(P \to t' : P), t : [\neg(P \to t' : P)] \vdash_{\mathsf{J4_{CS}}} \bot;$$

in other words, by soundness, the two hypotheses would not be $\mathsf{J4_{CS}}$-satisfiable. At the same time, the F-model constructed in the proof of Theorem 26 for the case of LP satisfies both of them. It remains to show that any $\mathsf{LP_{CS'}}$-model is also a $\mathsf{J4_{CS}}$-model, where CS′ stands for the maximal constant specification for LP that is not directly self-referential, whereas CS is the respective maximal constant specification for J4. $\mathsf{LP_{CS'}}$-models also require that R be transitive and $\mathcal{A}$ satisfy closure conditions C2, C3, C5, and Monotonicity. All axioms of J4 are also axioms of LP, and the definition of directly self-referential constants is logic independent, so $\mathsf{CS} \subset \mathsf{CS'}$. Thus,

[14] The idea to use this formula for K4 is due to Melvin Fitting.

the $*\mathsf{CS}'$-closure implies the $*\mathsf{CS}$-closure. So the supposedly unsatisfiable formulas have a model. This contradiction completes the proof. □

6 Non-Self-Referential Cases: D and K

In this section, we will show that $(\mathsf{JD}_{\mathsf{CS}})^\circ = \mathsf{D}$ and $(\mathsf{J}_{\mathsf{CS}})^\circ = \mathsf{K}$ for some non-self-referential constant specifications CS. Moreover, we will make sure that no realization of a modal theorem requires any self-referential cycles.

To construct such realizations, we divide both the set of constants and the set of justification variables into *levels* indexed by non-negative integers as follows. Let $\ell(c_i)$ and $\ell(x)$ denote the level of constant c_i and of variable x respectively. We require that consecutive constants from the same family have consecutive levels: $\ell(c_{i+1}) = \ell(c_i) + 1$. We also distribute constants and variables into levels in such a way that for each non-negative integer i both set

$$\{a_1 \mid a \text{ is a family of constants and } \ell(a_1) = i\}$$

and set

$$\{x \mid x \text{ is a justification variable and } \ell(x) = i\}$$

are infinite.

Let At be the set of all atomic justification terms: constants and variables, and let $\mathrm{At}(F)$ and $\mathrm{At}(t)$ denote the sets of all atomic terms that occur in formula F and in term t respectively. We extend the definition of level to terms and formulas as follows:

$$\ell(t) = \max\{\ell(p) \mid p \in \mathrm{At}(t)\}, \tag{19}$$

$$\ell(F) = \max\{\ell(p) \mid p \in \mathrm{At}(F)\}. \tag{20}$$

If $\mathrm{At}(F) = \emptyset$, we define $\ell(F) = 0$. For instance, $\ell(P) = 0$ for any sentence letter P. Let

$$\mathsf{CS} = \{c_n : c_{n-1} : \ldots : c_1 : A \in \mathsf{TCS}_{\mathsf{JL}} \mid \ell(c_1) > \ell(A)\} \tag{21}$$

for $\mathsf{JL} \in \{\mathsf{J}, \mathsf{JD}\}$. Such a constant specification is clearly axiomatically appropriate.

Theorem 29 *It is possible to realize* D *in* JD *and* K *in* J *without self-referentiality.*

Proof We reprove the Realization Theorem for D in $\mathsf{JD}_{\mathsf{CS}}$ and for K in J_{CS} for the respective CS from (21) making sure that whenever $t : F$ appears in the derivation of the realizing justification formula, $\ell(t) > \ell(F)$.

Since $\mathsf{JL}_{\mathsf{CS}} \subseteq \mathsf{JL}$, we have $(\mathsf{JD}_{\mathsf{CS}})^\circ \subseteq \mathsf{JD}^\circ = \mathsf{D}$ and $(\mathsf{J}_{\mathsf{CS}})^\circ \subseteq \mathsf{J}^\circ = \mathsf{K}$, so it remains to prove the other inclusion. Before doing so, we need to describe the behavior of $\Box$'s in cut-free Gentzen derivations for logics K and D.[15]

[15] It was suggested by Valentin Shehtman that this property is due to *uniformity* of these modal logics. For a discussion of uniform modal logics, see [6, 7].

We take Gentzen calculus G3c from [23] for classical propositional logic, i.e., we restrict the axioms to $\bot \Rightarrow$ and $P \Rightarrow P$ for sentence letters P.[16] The only modal rule to be added for logic K is

$$\frac{C_1, \ldots, C_n \Rightarrow B}{\Box C_1, \ldots, \Box C_n \Rightarrow \Box B}. \tag{22}$$

In addition, logic D enjoys

$$\frac{C_1, \ldots, C_n, D \Rightarrow}{\Box C_1, \ldots, \Box C_n, \Box D \Rightarrow}. \tag{23}$$

(Gentzen rules necessary for various modal logics can be found, for instance, in [9, 25]. The system for D seems to originate from [11]. See also [24] for the exposition of a syntactic cut-elimination for D, which subsumes the one for K since rule (22) is present in D.)

We define the *depth of an occurrence of* $\Box$ *in a modal formula* F by induction on the size of F: the outer $\Box$ in $\Box G$ has depth 0 in $\Box G$; for any occurrence of $\Box$ inside G, its depth in $\Box G$ is obtained by adding 1 to its depth in G.

We now define the *level of an occurrence of* $\Box$ *in a Gentzen derivation* as its depth in the formula it occurs in plus the number of modal rules (22) and (23) used on its branch after this occurrence.

All occurrences of $\Box$ in a cut-free Gentzen derivation can be divided into families of related occurrences. It is easy to prove that

Lemma 30 *In a Gentzen* K- *or* D*-derivation of* $\Rightarrow G$*, the levels of all occurrences of* $\Box$ *from a given family are equal to the depth of the family's occurrence in* G*.*

Thus, we can define the level of a family of $\Box$'s. Moreover, it is fairly obvious that all new $\Box$'s introduced by a particular instance of either (22) or (23) have the same level, which enables us to define the level of a given instance of a modal rule.

Let N be the largest level of $\Box$'s in a given cut-free derivation.

We use a proof of the Realization Theorem that transforms a given cut-free Gentzen derivation of a modal theorem G, i.e., of sequent $\Rightarrow G$, into a Hilbert derivation of its realization G^r by induction on the Gentzen derivation, whereby each sequent $\vdash \Gamma \Rightarrow \Delta$ is being transformed into $\Gamma^r \vdash \bigvee \Delta^r$.[17] A detailed description can be found in [2, 4, 5]. Our approach here is different in that we realize $\Box$'s according to their levels, which eventually enables us to avoid self-referentiality. We first describe the Realization Procedure along with level assignments. Then we show why self-referentiality does not occur.

[16] In G3c Weakening and Contraction rules are absorbed into the axioms. Here it is more convenient to have Weakening and Contraction present explicitly while keeping the axioms as plain as possible. This allows for a greater control over where $\Box$'s are introduced. It is important nevertheless that the systems we use be cut-free.

[17] As always, the empty disjunction is interpreted as $\bot$.

A cut-free derivation preserves the polarity of formulas, so we can divide families of $\Box$'s into positive and negative. We realize each negative family by a distinct justification variable of level $N-i$, where i is the level of this family of $\Box$'s. The same is done for the positive families that are introduced exclusively by Weakening. If at least one $\Box$ in a positive family is introduced by rule (22), such a family is realized by a sum of auxiliary variables $v_1+\cdots+v_l$, one variable per each use of (22) to introduce a $\Box$ from this family. (Note that rule (23) does not introduce new positive $\Box$'s.) These auxiliary variables are not real justification terms; they are placeholders to be replaced by actual terms in the course of the realization. The level of each auxiliary variable is also defined to be $N-i$, where i is the level of the respective instance of (22), or equivalently the level of the family of $\Box$'s the variable temporarily realizes. Let us call this *pre*liminary *realization* of $\Box$'s in the given Gentzen derivation a *prerealization*. As will be seen later, the prerealization is only changed by instances of modal rule (22).

The Gentzen axioms, propositional rules, and Contraction do not introduce any new $\Box$'s and can be translated from Gentzen into Hilbert using the standard propositional translation methods. Since the reasoning involved is purely propositional, there are no changes to the prerealization; in particular, no new terms or subformulas of type $t:F$ appear anywhere in the Hilbert derivation under construction due to Gentzen axioms or these rules.

Instances of Weakening can introduce formulas with $\Box$'s, so new terms (at least new to this Gentzen branch) may have to be introduced. The realization of $\Box$'s introduced by Weakening is done according to the prerealization, except that some auxiliary variables might have already been replaced by real justification terms during the translation of preceding modal rules. Since the reasoning needed to translate instances of Weakening is also purely propositional, no new terms are introduced, except those that realize new $\Box$'s, and no new subformulas of type $t:F$ appear in the Hilbert derivation, except those that realize new modal subformulas in the Gentzen derivation.

Thus, changes to the prerealization can happen only when instances of modal rules (22) and (23) are translated. To translate such instances, we use the Internalization Property (Lemma 9). This prompts an appearance of terms and formulas of type $t:F$ in the Hilbert derivation that do not themselves realize any $\Box$'s and modal formulas in the Gentzen derivation. Such terms may be subterms of realizing terms or, as in the case of rule (23), terms may simply disappear from the final justification formula and only remain present in the Hilbert derivation. Our goal is to show that self-referentiality does not occur even in such "hidden" terms and formulas.

Consider rule (22) first. By IH, we already have a Hilbert derivation of

$$C_1^r,\ldots,C_n^r\vdash B^r. \tag{24}$$

By Lemma 9, there exists a term $t(x_1,\ldots,x_n)$ such that

$$x_1:C_1^r,\ldots,x_n:C_n^r\vdash t(x_1,\ldots,x_n):B^r, \tag{25}$$

where each x_i is the prerealization of the negative $\Box$ in front of C_i in the conclusion of (22). Throughout the Hilbert proof, we substitute $t(x_1,\ldots,x_n)$ for the auxiliary

variable that corresponds to this instance of rule (22) in the sum realization of the family of the $\Box$ in front of B (in the conclusion of the modal rule). According to the proof of Lemma 9, each axiom A in derivation (24) gives rise to a constant c_1 in (25), to be taken from a fresh family of constants and used in $c_1:A$. Similarly, each use of the Axiom Internalization Rule, $c_k:c_{k-1}:\ldots:c_1:A$, in (24) requires a new constant c_{k+1}, to be used in $c_{k+1}:c_k:c_{k-1}:\ldots:c_1:A$ in (25), where naturally c_{k+1} is taken from the same family of constants as $c_1,\ldots,c_k$. In the latter case, the level of c_{k+1} is predetermined by the level of c_k. In the former case, we choose a new constant so that $\ell(c_1)=N-i$, where i is the level of this instance of rule (22), or equivalently the level of the family of the $\Box$ in front of B. Each hypothesis C_i^r in (24) acquires in (25) variable x_i according to the prerealization. The level of these variables is also $N-i$, where i is the level of this instance of (22). All the new constants and variables become subterms of term $t(x_1,\ldots,x_n)$, which realizes the $\Box$ in front of B.

Rule (23) is treated similarly. Here the Internalization yields

$$x_1:C_1^r,\ldots,x_n:C_n^r,y:D^r\vdash t(x_1,\ldots,x_n,y):\bot,$$

from which $\bot$ can be easily derived by means of axiom A7, $t(x_1,\ldots,x_n,y):\bot\to\bot$, and *modus ponens*. We use the same guidelines for assigning levels to new constants c_1 inside $t(x_1,\ldots,x_n,y)$ as the ones we used for rule (22) except that there is no B and hence no $\Box$ to be realized. But we can still use the level of the instance of (23) to figure out the level of the new constants. Note that no change to the prerealization happens here: term $t(x_1,\ldots,x_n,y)$ simply disappears even though the constants that comprise it remain in the Hilbert derivation. These are the "hidden" constants referred to earlier. These constants, or rather constants from the same family, can later be incorporated into terms to replace positive $\Box$'s in the subsequent instances of rule (22).

We have chosen the level of each variable and auxiliary variable in the prerealization to be equal to $N-i$, where i is the level of the family of $\Box$'s realized by the respective (auxiliary) variable. We have made sure that this correspondence remains valid for constants c_1 introduced in the modal rules if those constants are used in the realization of some positive $\Box$. In instances of rule (23), the level has been matched to that of the rule. It remains to verify that constants c_{k+1} introduced during Internalizations also comply. It will then follow that the substitutions of a term for an auxiliary variable in translating instances of (22) do not violate this preset harmony. These substitutions are the source of self-referentiality in stronger modal logics.

Let us prove that the level of c_{k+1} still matches the level of the rule that prompts its introduction. Indeed, as we have seen, constants are only introduced to be used in Axiom Internalizations. Each instance of Axiom Internalization remains in the Hilbert derivation throughout the subsequent propositional Gentzen steps until an appearance of the next instance of a modal rule because the translation of a propositional Gentzen rule, logical or structural alike, only appends the existing Hilbert derivation. Each constant c_1 has level $N-i$ by construction, where i is the level of the corresponding instance of the modal rule. Suppose constants c_k also satisfy this property. Consider an instance of a modal rule whose translation has introduced constant c_{k+1}.

It gets introduced because c_k had already been used in the Hilbert derivation. By IH, the level of c_k is equal to $N - i$, where i is the level of the instance of the rule that introduces c_k. (Note that every constant is introduced only once because we always choose a fresh constant and substitutions do not influence the introduction of constants, only of the formulas justified by these constants.) It is easy to observe that if instance I2 of a modal rule follows instance I1 of a modal rule in a Gentzen derivation with no other modal rules on the branch between them, then $\ell(\text{I2}) = \ell(\text{I1}) - 1$. Thus, $\ell(c_{k+1}) = \ell(c_k) + 1 = N - \ell(\text{I1}) + 1 = N - \ell(\text{I2})$.

Therefore, whenever a term t replaces an auxiliary variable, this term consists entirely of constants and variables whose level is $N - i$, where i is the level of the family of $\Box$'s realized by t. The replaced auxiliary variable has the exact same level, so substitutions do not change the level of any terms or formulas.

We now prove that whenever $t:F$ appears in the translation, $\ell(t) > \ell(F)$ by induction on the depth of the Gentzen derivation. As discussed before, we only need to consider Gentzen steps corresponding to modal rules. Consider an instance I of a modal rule of level i. All modal rules in the subtree whose root is the premise of I and all $\Box$'s present in this subtree have levels strictly greater than i. Therefore, all variables and constants in the Hilbert derivation of the realization of the premise of I have levels strictly smaller than $N - i$ (here we consider the final form of the derivation after substitutions have replaced all auxiliary variables). So all the formulas in the Hilbert derivation before Internalization have levels $< N - i$. All the constants and variables introduced during Internalization have level $N - i$, and so do all the new terms constructed by Internalization. Therefore, whenever formula $s:F$ appears in the internalized derivation, i.e., the derivation of the conclusion of I, $\ell(s) = N - i > \ell(F)$ because F was present in the derivation before Internalization (see proof of Lemma 9).

Demonstrating that self-referentiality does not occur is now easy. Suppose formulas

$$t_2:F_1(t_1), \quad \ldots, \quad t_m:F_{m-1}(t_{m-1}), \quad t_1:F_m(t_m)$$

are present in the final Hilbert derivation. That would imply that

$$\begin{cases} \ell(t_2) > \ell(F_1(t_1)) \geq \ell(t_1), \\ \vdots \\ \ell(t_m) > \ell(F_{m-1}(t_{m-1})) \geq \ell(t_{m-1}), \\ \ell(t_1) > \ell(F_m(t_m)) \geq \ell(t_m). \end{cases}$$

In other words, we would have

$$\ell(t_m) > \ell(t_{m-1}) > \cdots > \ell(t_2) > \ell(t_1) > \ell(t_m),$$

which is impossible.

We have shown that self-referentiality can be avoided in formulas that realize all modal theorems of K and D, as well as in derivations of these realizing formulas. $\Box$

7 What Is Wrong with Negative Introspection?

Modal logics with negative introspection K5, K45, KD45, and S5 also have their justification counterparts. These counterparts are obtained by adding a new unary operation ? on justification terms. The role of ? with respect to negative introspection is similar to that of ! with respect to positive introspection. We discuss the difficulties in expanding the study of self-referentiality to these logics using JT45, the counterpart of S5, as a representative example.

JT45 is obtained from LP by adding

A6. *Negative Introspection Axiom* $\neg t:F \rightarrow ?t:\neg t:F$

(see [3, 20, 21]). It would seem that the argument from Example 1, which only requires T reasoning, can therefore be performed in S5 equally well. This hints at the self-referentiality of S5, which should be provable through terms of JT45.

However, this program is far from completion. Our method for demonstrating self-referentiality for S4 and several other logics involves showing that the only way to falsify $\neg t:[\neg(P \rightarrow t':P)]$ is by using directly self-referential constants. On the face of it, this can be done in two ways. The first option is to analyze all possible derivations to show that they all feature self-referential constants. But studying the properties of Hilbert derivations is an unwieldy task given an absence of the subformula property. So we resort to showing the contrapositive: we eliminate self-referential constants and in their absence are able to construct a countermodel. In doing so we rely on the existence of a minimal admissible evidence function such that for some world w

$$w \in \mathcal{A}(t, \neg(P \rightarrow t':P)) \quad \text{but} \quad w \notin \mathcal{A}(t', P).$$

Logic JT45 also has an epistemic semantics complete with admissible evidence functions. They have to satisfy an additional closure condition

C5. $[\mathcal{A}(t, F)]^c \subseteq \mathcal{A}(?t,\ \neg t:F)$,

where $[\cdot]^c$ means the complement within W. Unfortunately, all variants of F-models for JT45 feature another requirement that proves to be non-recursive. Pacuit in [20] imposes a condition that he calls *Negative Proof Checker*, which states: "If there is a v such that wRv and $\mathcal{M}, v \Vdash \neg F$ [...], then $w \in \mathcal{A}(?t, \neg t:F)$." This condition is non-constructive since $\mathcal{A}(?t, \cdot)$ is not determined solely by $\mathcal{A}(t, \cdot)$. It is not clear, in particular, how to satisfy this condition if F contains $?t$. Rubtsova in [21] and Artemov in [3] replace this requirement by the so-called *Strong Evidence* property: "if $w \in \mathcal{A}(t, F)$, then $\mathcal{M}, w \Vdash t:F$."[18] Although this condition looks a little nicer, it is in fact equivalent to Pacuit's condition and thus is as hard to satisfy. It also ties a statement about admissible evidence $w \in \mathcal{A}(t, F)$ to the truth of formula F at worlds accessible from w, which is implied by $w \Vdash t:F$, thus violating the useful separation of truth from admissible evidence in LP and weaker logics: the truth of formulas there does not affect the admissibility of evidence.

[18]This formulation is from [3]; Rubtsova's condition is almost identical. The term *strong evidence* is due to Fitting and is a property of the canonical models for all justification logics.

Still, a condition of such type is necessary for the validity of the Negative Introspection Axiom. Indeed, $\neg t:F$ could be true simply because F is false in some accessible world, which has nothing to do with whether t is admissible evidence. But to validate $?t:\neg t:F$ we must make $?t$ admissible evidence for $\neg t:F$. It seems that a different semantics is needed to handle justification logics with negative introspection.

8 Future Research

There are many directions in which this study can be developed.

Self-referentiality results can be used to prove structural properties of Gentzen modal derivations, e.g., the unavoidability of double introduction of the same family of $\Box$'s on the same branch for directly self-referential modal logics. This opens new applications of Justification Logic to structural proof theory.

It remains to see what triggers self-referentiality. It appears that self-referentiality is tied to the possibility of mixing levels of $\Box$'s in a Gentzen derivation, to the non-uniformity of a modal logic, but we need a larger sample set to make any definite conclusions. We conjecture that the statement of Lemma 30 can be viewed as a purely modal formulation of a sufficient criterion for non-self-referentiality. It would be interesting to see whether it is also necessary.

We still do not know of an example when self-referentiality is required but direct self-referentiality can be avoided.

A manageable semantics for logics with negative introspection could open a lot of avenues into their study, including self-referentiality, decidability, complexity, etc.

Another direction is a deeper study of self-referentiality where it is unavoidable.[19] Some modal theorems in, say, S4 can be realized without self-referentiality, e.g., all theorems of D. What is the non-self-referential fragment of S4 and of other modal logics? Are such fragments decidable? Do they have a nice axiomatization? Can these fragments be constructed uniformly for different modal logics?

Acknowledgements The author is greatly indebted to Sergei Artemov, Lev Beklemishev, Melvin Fitting, Vladimir Krupski, and Valentin Shehtman, whose advice has helped to shape this paper at various stages of its creation. The author thanks the anonymous referees for careful reading and useful suggestions.

References

1. Artemov, S.N.: Operational modal logic. Technical Report MSI 95–29, Cornell University, December 1995
2. Artemov, S.N.: Explicit provability and constructive semantics. Bull. Symb. Log. **7**(1), 1–36 (2001)
3. Artemov, S.N.: The logic of justification. Technical Report TR-2008010, CUNY Ph.D. Program in Computer Science, September 2008
4. Brezhnev, V.N., Kuznets, R.: Making knowledge explicit: How hard it is. Theor. Comput. Sci. **357**(1–3), 23–34 (2006)

[19] This direction of research was suggested by Vladimir Krupski.

5. Brezhnev, V.N.: On explicit counterparts of modal logics. Technical Report CFIS 2000–05, Cornell University, 2000
6. Chagrov, A., Zakharyaschev, M.: Modal Logic. Oxford Logic Guides, vol. 35. Oxford University Press, Oxford (1997)
7. Fine, K.: Normal forms in modal logic. Notre Dame J. Form. Log. **16**(2), 229–237 (1975)
8. Fitting, M.: The logic of proofs, semantically. Ann. Pure Appl. Log. **132**(1), 1–25 (2005)
9. Fitting, M.: Modal proof theory. In: Blackburn, P., van Benthem, J., Wolter, F. (eds.) Handbook of Modal Logic. Studies in Logic and Practical Reasoning, vol. 3, pp. 85–138. Elsevier, Amsterdam (2007)
10. Gettier, E.L.: Is justified true belief knowledge? Analysis **23**(6), 121–123 (1963)
11. Goble, L.F.: Gentzen systems for modal logic. Notre Dame J. Form. Log. **15**(3), 455–461 (1974)
12. Goldman, A.I.: A causal theory of knowing. J. Philos. **64**(12), 357–372 (1967)
13. Hendricks, V.F.: Active agents. J. Log. Lang. Inform. **12**(4), 469–495 (2003)
14. Krupski, N.V.: On the complexity of the reflected logic of proofs. Theor. Comput. Sci. **357**(1–3), 136–142 (2006)
15. Kuznets, R.: On self-referentiality in modal logic. In: 2005–06 Winter Meeting of the Association for Symbolic Logic, The Hilton New York Hotel, New York, NY, December 27–29, 2005. Bull. Symb. Log., vol. 12(3), p. 510. Association for Symbolic Logic, September 2006
16. Kuznets, R.: Complexity issues in justification logic. PhD thesis, CUNY Graduate Center, May 2008
17. Kuznets, R.: Self-referentiality of justified knowledge. In: Hirsch, E.A., Razborov, A.A., Semenov, A., Slissenko, A. (eds.) Proceedings of the Third International Computer Science Symposium in Russia, CSR 2008, Moscow, Russia, June 7–12, 2008. Lecture Notes in Computer Science, vol. 5010, pp. 228–239. Springer, Berlin (2008)
18. Lehrer, K., Paxson, T. Jr.: Knowledge: Undefeated justified true belief. J. Philos. **66**(8), 225–237 (1969)
19. Mkrtychev, A.: Models for the logic of proofs. In: Adian, S., Nerode, A. (eds.) Proceedings of the 4th International Symposium Logical Foundations of Computer Science, LFCS'97, Yaroslavl, Russia, July 6–12, 1997. Lecture Notes in Computer Science, vol. 1234, pp. 266–275. Springer, Berlin (1997)
20. Pacuit, E.: A note on some explicit modal logics. In: Proceedings of the 5th Panhellenic Logic Symposium, Athens, Greece, July 25–28, 2005, pp. 117–125. University of Athens, Athens (2005)
21. Rubtsova, N.: Evidence reconstruction of epistemic modal logic S5. In: Grigoriev, D., Harrison, J., Hirsch, E.A. (eds.) Proceedings of the First International Computer Science Symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8–12, 2006. Lecture Notes in Computer Science, vol. 3967, pp. 313–321. Springer, Berlin (2006)
22. Smoryński, C.: Self-Reference and Modal Logic. Universitext. Springer, Berlin (1985)
23. Troelstra, A.S., Schwichtenberg, H.: Basic Proof Theory, 2nd edn. Cambridge Tracts in Theoretical Computer Science, vol. 43. Cambridge University Press, Cambridge (2000)
24. Valentini, S.: The sequent calculus for the modal logic D. Boll. Unione Mat. Ital. Sez. A **7**, 455–460 (1993)
25. Wansing, H.: Sequent calculi for normal modal propositional logics. J. Log. Comput. **4**(2), 125–142 (1994)