Justification Awareness Models

Sergei Artemov(✉)

The City University of New York, The Graduate Center,
365 Fifth Avenue, New York City, NY 10016, USA
sartemov@gc.cuny.edu

Abstract. Justification Awareness Models, *JAM*s, incorporate two principal ideas: (i) *justifications are prime objects of the model*: knowledge and belief are defined evidence-based concepts; (ii) *awareness restrictions are applied to justifications* rather than to propositions, which allows for the maintaining of desirable closure properties. *JAM*s naturally include major justification models, Kripke models and, in addition, represent situations with multiple possibly fallible justifications. As an example, we build a *JAM* for Russell's well-known Prime Minister scenario which, in full generality, was previously off the scope of rigorous epistemic modeling.

Keywords: Modal logic · Justification logic · Epistemology
Knowledge · Belief

1 Context and Motivations

Proof systems of justification logic and general purpose classes of models for these systems have been studied in [1–3,9,10,16,18,20] and many other sources. However, for formalizing epistemic scenarios, one needs specific *domain-dependent models* with additional features that are not necessary for standard soundness and completeness analysis of proof systems.

Awareness is an important concept in epistemic modeling, but, when applied to propositions directly, it may seriously diverge from the intuition due to lack of natural closure properties [7,8,17]. We suggest applying awareness to justifications

$$agent\ is\ aware/unaware\ of\ a\ justification\ t\ for\ a\ proposition\ F$$

rather then to propositions "agent is aware/unaware of a proposition F"; this approach allows for the maintaining of natural closure properties.

We introduce *justification awareness models*, *JAM*s, in which justifications are primary objects and a distinction is made between *accepted* and *knowledge-producing* justifications. In *JAM*s, belief and knowledge are derived notions which depend on the status of supporting justifications. We argue that *JAM*s can work in situations in which standard non-hyperintensional tools (Kripke, topological, algebraic) fail to fairly represent the corresponding epistemic structure.

S. Artemov and A. Nerode (Eds.): LFCS 2018, LNCS 10703, pp. 22–36, 2018.
https://doi.org/10.1007/978-3-319-72056-2_2

2 Preliminaries

Standard modal epistemic models have "propositional" precision, i.e., they do not distinguish sentences with the same truth values at each possible world. The expressive power of such models for analysis of justification, belief, and knowledge is rather limited, and so we have to "go hyperintensional."[1] Specifically, if, at all possible worlds, t is a justification for F

$$\Vdash t{:}F,$$

and G has the same truth value as F

$$\Vdash F \leftrightarrow G,$$

we still cannot conclude that t is a justification for G

$$\nVdash t{:}G.$$

A natural example from mathematics: both statements $0 = 0$ and *Fermat's Last Theorem*, FLT, are true (proven) mathematical facts and hence are true at all possible worlds. However, we cannot claim that a proof of $0 = 0$ is a proof of FLT as well.

A sample justification logic analysis of some standard epistemic situations (Gettier examples, Red Barn example) is presented in [2] using justification Fitting models [9] though, due to the relative simplicity of those examples, this analysis could be replicated in a bi-modal language (cf. [21]).

However, we cannot go much farther without adopting a justification framework: the situation changes when we have to represent several conflicting pieces of evidence for a stated fact, cf. the following Russell example of 1912 ([19]):

> *If a man believes that the late Prime Minister's last name began with a 'B,' he believes what is true, since the late Prime Minister was Sir Henry Campbell Bannerman*[2]*. But if he believes that Mr. Balfour was the late Prime Minister, he will still believe that the late Prime Minister's last name began with a 'B,' yet this belief, though true, would not be thought to constitute knowledge.*

To keep it simple, we consider proposition Q

the late Prime Minister's last name began with a 'B,'

with two justifications for Q: the right one r and the wrong one w; the agent chooses w as a reason to believe that Q holds.

To avoid a misleading reduction of failures of justifications to "false premises," consider another Russell example from [19].

[1] From [6]: "Hyperintensional contexts are simply contexts which do not respect logical equivalence".

[2] Which was true in 1912.

If I know that all Greeks are men and that Socrates was a man, and I infer that Socrates was a Greek, I cannot be said to-know-that Socrates was a Greek, because, although my premisses and my conclusion are true, the conclusion does not follow from the premisses.

This Russell's example illustrates that "false premises" in the Prime Minister story is an instance of a more general phenomenon: an erroneous justification which, in principle, can fail for many different reasons: unreliable premises, hidden assumptions, deduction errors, an erroneous identification of the goal sentence, etc.[3]

There is a mathematical version of the story with a true proposition and its two justifications; one is correct, the other is not.

Consider the picture[4]*:*

$$\frac{1\cancel{6}}{\cancel{6}4} = \frac{1}{4}. \tag{1}$$

The true proposition is "16/64 = 1/4," the right justification is dividing both the numerator and the denominator by 16, and the wrong (but shorter and more attractive) justification is simplifying as in (1).

Given these considerations, we prefer speaking about *erroneous justifications* in a general setting without reducing them to propositional entities such as "false premises." To be specific, we'll continue with Russell's Prime Minister example.

To formalize Russell's scenario in modal logic (cf. [21]), we introduce two modalities: **K** for knowledge and **J** for justified belief. In the real world,

– Q holds;
– $\mathbf{J}Q$ holds, since the agent has a justification w for Q;
– $\mathbf{K}Q$ does not hold;

thus yielding the set of assumptions

$$\Gamma = \{Q,\ \mathbf{J}Q\ \neg\mathbf{K}Q\}.$$

However, Γ doesn't do justice to Russell's scenario: the right justification r is not represented and Γ rather corresponds to the same scenario but lacking r. The epistemic structure of the example is not respected.

Within the *JAM* framework, we provide a model for Russell's Prime Minister example which, we wish to think, fairly represents its intrinsic epistemic structure.

[3] Moreover, one can easily imagine knowledge-producing reasoning from a source with false beliefs (both an atheist and a religious scientist can produce reliable knowledge products though one of them has false beliefs), so "false premises" are neither necessary nor sufficient for a justification to fail.

[4] Which the author saw on the door of the Mathematics Support Center at Cornell in 2017.

3 Generic Logical Semantics of Justifications

What kinds of logical objects are justifications? When asked in a mathematical context "what is a predicate?" we have a ready answer: a subset of a Cartesian product of the domain set. Within an exact mathematical theory, there should be a similar kind of answer to the question "what is a justification?"

We consider this question in its full generality which, surprisingly, yields a clean and meaningful answer. We assume the language of justification logic consists of two disjoint sets of syntactic objects:

1. a set of **justification terms** Tm;
2. a set of **formulas** Fm, built inductively from propositional atoms using Boolean connectives and the justification formula formation rule: if F is a formula, $F \in Fm$, and t a justification term, $t \in Tm$, then $t{:}F$ is again a formula, $t{:}F \in Fm$.

The meaning assigned to formulas is a classical truth value, 0 for *false* and 1 for *true*, and we retain classical logic behavior for propositional connectives. The key item is to give meaning to justification terms, and this will be a *set of formulas* interpreted as *the set of formulas for which it is a justification.* A formal definition follows.

Definition 1 (Basic Model). *A basic model, simply called $*$, consists of an interpretation of the members of Fm, and an interpretation of the members of Tm.*

The interpretation of a formula in a basic model is a truth value. That is,

$$* : Fm \mapsto \{0,1\}.$$

We assume the Boolean truth tables: $(X \rightarrow Y)^* = 1$ if and only if $X^* = 0$ or $Y^* = 1$, etc. Let also $\models_* X$ stand for $X^* = 1$.

We interpret justification terms as sets of formulas. That is,

$$* : Tm \mapsto 2^{Fm}.$$

Our final requirement connects the two mapping roles that $*$ plays in a basic model. For any $X \in Fm$ and any $t \in Tm$,

$$\models_* t{:}X \text{ if and only if } X \in t^*.$$

It is easy to check that any mapping $*$ from propositional letters to truth values, and from justification terms to sets of formulas, determines a unique basic model.

So far, a basic model is merely a classical propositional model in which justification assertions $t{:}F$ are treated as *independent propositional atoms.*

Note that while propositions are interpreted semantically as truth values, justifications are interpreted syntactically as sets of formulas. This is a principal *hyperintensional* feature: a basic model may treat distinct formulas F and G as equal, i.e. $F^* = G^*$, but still be able to distinguish justification assertions $t{:}F$ and $t{:}G$, e.g., when $F \in t^*$, but $G \notin t^*$ yielding $\models_* t{:}F$ but $\not\models_* t{:}G$.

Definition 2. *Let S a set of formulas, $S \subseteq Fm$, and X be a formula, $X \in Fm$. We write $S \vdash X$ if X is derivable from S in classical logic that treats justification assertions $t{:}F$ as propositional atoms (with Modus Ponens as the only rule of inference). We say that S is consistent if $S \nvdash \bot$.*

A basic model of S is merely a possible world containing S in the canonical model, i.e., a maximal consistent set Γ of formulas, with the convenience agreement reading $t{:}F \in \Gamma$ as $F \in \{X \mid t{:}X \in \Gamma\}$. In this respect, basic models and the canonical model are slightly different but obviously equivalent ways of presenting the same object. When we move to more sophisticated models (Fitting models, modular models), the advantage of dealing with sets and operations (e.g. basic models) over logical conditions (e.g. the canonical model) becomes clear.

Definition 3. *For $S \subseteq Fm$, $BM(S)$ is the class of all basic models of S.*

Theorem 1. *Each set of formulas S is sound and complete with respect to its class of basic models $BM(S)$. In other words, $S \vdash F$ iff F is true in each basic model of S.*

Proof. This theorem is merely a reformulation of the soundness and completeness of classical propositional logic with hypotheses. Indeed, if $S \vdash F$ and $\models_* S$, then $\models_* F$ since propositional derivations respect validity.

If $S \nvdash F$, then there is a Boolean evaluation $*$ which makes all formulas from S true, $S^* = 1$, and F false, $F^* = 0$. In this case, there are two types of atomic propositions: propositional letters P and justification assertions $t{:}X$. Define

$$t^* = \{X \mid (t{:}X)^* = 1\}$$

and note that $(t{:}X)^* = 1$ iff $X \in t^*$. Therefore, $*$ is a propositional evaluation and $*$ is a basic model yielding the same truth values of atomic formulas P and $t{:}X$. Since $S^* = 1$ and $F^* = 0$, we have $\models_* S$ and $\not\models_* F$ for basic model $*$.

An easy corollary: $\vdash F$ iff F is a tautology (with $t{:}X$es as distinct propositional atoms).

Example 1. In Definition 2, take $S = \emptyset$.

1. For any justification term t,
$$\nvdash t{:}F.$$
Straightforward, since $t{:}F$ is not a propositional tautology. For a specific countermodel, define $t^* = \emptyset$ for each term $t \in Tm$, which makes $\not\models_* t{:}F$.
2. For any propositional letter P, and term t,
$$\nvdash t{:}P \to P.$$
Likewise, this holds because $t{:}P \to P$ is not a propositional tautology. Specifically, put $t^* = Fm$ and $P^* = 0$, with other assignments being arbitrary. In this model, all justification assertions are true, but $t{:}P \to P$ is false.

3. For any propositional letter P, and term t,

$$\not\vdash P \to t{:}P.$$

   Again, this holds since $P \to t{:}P$ is not a propositional tautology. For example, put $t^* = \emptyset$ and $P^* = 1$. In this model, t is not a justification for P (i.e., $\not\models_* t{:}P$) and $P \to t{:}P$ is false.
4. A somewhat less trivial example illustrating hyperintensionality: for a justification variable x and formula F

$$\not\vdash x{:}F \to x{:}(F \wedge F).$$

   A high-level argument is the same: formulas $x{:}F$ and $x{:}(F \wedge F)$, evaluated from a Boolean point of view, can be regarded as distinct propositional variables. Hence $x{:}F \to x{:}(F \wedge F)$ is not a tautology. For a countermodel, take $x^* = \{F\}$. Then $\models_* x{:}F$, but $\not\models_* x{:}(F \wedge F)$. This demonstrates hyperintensionality of a justification logic base, since F and $F \wedge F$ are provably equivalent, but not $x{:}F$ and $x{:}(F \wedge F)$.

4 Basic Justification Logic J^-

Within the Justification Logic framework, there are two sorts of logical objects: justification terms *Tm* and formulas *Fm*. Let us become more specific about both.

- For *Tm*, reserve a set of justification constants $a, b, c, \ldots$ with indices, and variables $x, y, z, \ldots$ with indices. Justification terms are built from constants and variables by a binary operation $\cdot$ (application).
- Formulas are built from propositional letters $p, q, r, \ldots$ (with indices) and Boolean constant $\bot$ (falsum) by the standard Boolean connectives $\wedge, \vee, \to, \neg$ with a new formation rule: *whenever t is a justification term and F is a formula, $t : F$ is a formula (with the informal reading "t is a justification for F").* For better readability, we will interchangeably use brackets $0, 0$ and parentheses $(,)$. Our preferred notation is $[s \cdot t]{:}(F \to G)$ which is the same as $(s \cdot t){:}(F \to G)$.

The *logical system* J^- consists of two groups of postulates.

- **Background logic**: axioms of classical propositional logic, rule *Modus Ponens*.
- **Application**: $s{:}(F \to G) \to (t{:}F \to [s \cdot t]{:}G)$.

Basic models corresponding to J^- are those in which the application axiom holds. They can be specified by a natural combinatorial condition.

Definition 4. *For sets of formulas S and T, we define*

$$S \rhd T = \{F \mid G \to F \in S \text{ and } G \in T \text{ for some } G\}.$$

Informally, $S \rhd T$ is the result of applying Modus Ponens once to all members of S and of T (in a given order).

Theorem 2. *$BM(\mathsf{J}^-)$ is the class of basic models with the following closure condition*

$$s^* \triangleright t^* \subseteq [s{\cdot}t]^*. \qquad (2)$$

Proof. Let us assume the closure condition (2) and check the validity of the application axiom. Indeed, $\models_* s{:}(F \to G)$ and $\models_* t{:}F$ yield $(F \to G) \in s^*$ and $F \in t^*$. By the closure condition, $G \in [s{\cdot}t]^*$, i.e., $\models_* [s{\cdot}t]{:}G$.

Now assume the application axiom and derive the closure condition (2). Let $(F \to G) \in s^*$ and $F \in t^*$. By definition, this yields $\models_* s{:}(F \to G)$ and $\models_* t{:}F$. By the application axiom, $\models_* [s{\cdot}t]{:}G$, hence $G \in [s{\cdot}t]^*$.

Example 2. None of the formulas from Example 1: $t:F$, $t:P \to P$, $P \to t:P$, $x:F \to x:(F \wedge F)$ is derivable in J^-. Indeed, every specific evaluation from Example 1.1–3 satisfies the closure condition (2), hence their countermodels are J^--models. Consider the latter formula 4. Put $x^* = \{F\}$ and $t^* = Fm$ for all other terms t. The closure condition (2) holds vacuously, hence $*$ is a J^--model. Obviously, $\models_* x{:}F$ and $\not\models_* x{:}(F \wedge F)$.

Constants in justification logic are used to denote justifications of assumptions, in particular, axioms. Indeed, as we have already seen in Example 2, no formula $t{:}F$ is derivable in J^-. In particular, no logical axiom is assumed justified in J^- which is not realistic.

Definition 5. *A set X of formulas is reflexive if for each $s{:}t{:}F \in X$, $t{:}F$ is also in X. By constant specification CS we understand a reflexive set of formulas of the type*

$$c_n{:}c_{n-1}{:}c_{n-2}{:} \ \ldots \ c_1{:}A$$

where A is a J^--axiom and c_i are justification constants. The major classes of constant specifications are empty, total— (each constant is a justification for each axiom), axiomatically appropriate (each axiom has a justification at any depth).

Let CS be a constant specification. Then by $\mathsf{J}^-(CS)$, we understand J^- with additional axioms CS. A *CS-model* is a model in which all formulas from CS hold.

Corollary 1. *Basic models for $\mathsf{J}^-(CS)$ are the basic CS-models for J^-. $\mathsf{J}^-(CS)$ is sound and complete with respect to the class of its basic models.*

4.1 Other Justification Logics

There is a whole family of justification logics and they all extend J^-; the reader is referred to [2,11] for details. Here we list just the main systems of justification logic for purposes of general orientation.

Logic J is obtained from J^- by adding a new operation on justifications '+' and the principle

$$s{:}F \vee t{:}F \to [s+t]{:}F.$$

Logics JD, JT, $\mathsf{J4}$, $\mathsf{J5}$, etc., are obtained by adding the corresponding combination of principles

$$D = \neg t{:}\bot,$$

$$T = t{:}F \rightarrow F,$$

$$4 = t{:}F \rightarrow !t{:}t{:}F,$$

$$5 = \neg t{:}F \rightarrow ?t{:}\neg t{:}F.$$

The family of justification logics has now grown to be infinite, cf. [11].

4.2 Sharp Models

In closure condition (2) from Theorem 2, one cannot, generally speaking, replace the inclusion "$\subseteq$" by the equality "$=$" without violating completeness Theorem 1.

Indeed, fix a justification constant 0 and consider logic

$$\mathcal{L} = \mathsf{J}^{-} + \{\neg 0{:}F \mid F \in Fm\}.$$

Informally, justification 0 receives empty evaluation in any basic model, $0^* = \emptyset$. We claim that formula $G = \neg[0\cdot 0]{:}P$ is not derivable in $\mathcal{L}$, but is true in any basic model of $\mathcal{L}$ with the closure condition $s^* \triangleright t^* = [s\cdot t]^*$. To show that $\mathcal{L} \not\vdash G$, it suffices to find a basic model for $\mathcal{L}$ in which G is false. Consider a basic model $\sharp$ such that $0^\sharp = \emptyset$ and $t^\sharp = Fm$ for any other justification term t. Obviously, the closure condition from Theorem 2, together with $0^\sharp = \emptyset$, is met. Therefore, $\sharp$ is a basic model of $\mathcal{L}$. It is immediate that G is false in $\sharp$, since $[0\cdot 0]^\sharp = Fm$. On the other hand, G holds in any basic model of $\mathcal{L}$ with the closure condition $[0\cdot 0]^* = 0^* \triangleright 0^*$. Indeed, in such a model, $[0\cdot 0]^* = \emptyset$ since $0^* = \emptyset$ and $\emptyset \triangleright \emptyset = \emptyset$.

Definition 6. Sharp *basic models are those in which the application closure condition has the form*

$$[s\cdot t]^* = s^* \triangleright t^*. \tag{3}$$

Note that a sharp model is completely defined by evaluations of atomic propositions and atomic justifications.

5 Justification Awareness

We need more expressive power to capture epistemic differences between justifications and their use by the knower. Some justifications are knowledge-producing, some are not. The agent makes choices on which justifications to base an agent's beliefs/knowledge and which justifications to ignore in this respect. These actions are present in epistemic scenarios, from which we will primarily focus on Russell's Prime Minister example, which has them all:

- there are justifications w (Balfour was the late prime minister) and r (Bannerman was the late prime minister) for Q;
- r is knowledge-producing whereas w is not;
- the agent opts to base his belief on w and ignores r;
- the resulting belief is evidence-based, but is not knowledge.

5.1 Justification Awareness Models

Fix $\mathsf{J}^-(CS)$ for some axiomatically appropriate constant specification CS.

Definition 7. *A set X of justification terms is properly closed if X contains all constants and is closed under applications. If X is a set of justification terms, then by $\overline{X}$ we mean the proper closure of X, i.e., the minimal properly closed superset of X.*

Definition 8. *A (basic) Justification Awareness Model is $(*, \mathcal{A}, \mathcal{E})$ where*

- *$*$ is a basic $\mathsf{J}^-(CS)$-model;*
- *$\mathcal{A} \subseteq Tm$ is a properly closed set $\mathcal{A}$ of accepted justifications;*
- *$\mathcal{E} \subseteq Tm$ is a properly closed set $\mathcal{E}$ of knowledge-producing justifications.*

Unless stated otherwise, we also assume consistency of accepted justifications: $\models_* \neg t{:}\bot$ for any $t \in \mathcal{A}$, and factivity of knowledge-producing justifications, $\models_* t{:}F \rightarrow F$ for each F and each $t \in \mathcal{E}$. In models concerning beliefs rather then knowledge, the component $\mathcal{E}$ can be dropped.

Both sets $\mathcal{A}$ and $\mathcal{E}$ contain all constants. This definition presumes that constants in a model are knowledge-producing and accepted.

Definition 9. *In a JAM $(*, \mathcal{A}, \mathcal{E})$, a sentence F is believed if there is $t \in \mathcal{A}$ such that $\models_* t{:}F$. Sentence F is known if there is $t \in \mathcal{A} \cap \mathcal{E}$ such that $\models_* t{:}F$.*

By *ground term* we understand a term containing no (justification) variables. In other words, a term is ground iff it is built from justification constants only.

Sets of accepted and knowledge-producing justifications overlap on ground terms but otherwise can be in a general position[5]. There may be accepted, but not knowledge-producing, justifications and vice versa. So, *JAM*s do not analyze **why** certain justifications are knowledge-producing or accepted, but rather provide a formal framework that accommodates these notions.

5.2 Single-Conclusion Justifications

The notions of *accepted* and *knowledge-producing* justifications should be utilized with some caution. Imagine a justification t for F (i.e., $t{:}F$ holds) and for G ($t{:}G$) such that, intuitively, t is a knowledge-producing justification for F but not for G. Is such a t knowledge-producing, trustworthy, acceptable for a reasonable agent? The answers to these questions seem to depend on F and G, and if we prefer to handle justifications as objects rather than as justification assertions, it is technically convenient to assume that justifications are *single-conclusion* (or, equivalently, *pointed*):

$$\textit{there is at most one formula } F \textit{ such that } t{:}F \textit{ holds.}$$

[5] In principle, one could consider smaller sets $\mathcal{A}$, which would correspond to the high level of skepticism of an agent who does not necessarily accept logical truths (axioms) as justified. We leave this possibility for further studies.

Conceptually, by turning to pointed justifications, one does not lose generality: if p is a proof of F and of something else, then the same p with a designated statement F, symbolically, a pair (p, F), can be regarded as a single-conclusion (or pointed) proof of F.

In model $\mathcal{R}$ for the Russell Prime Minister example, Sect. 6, all justifications are pointed.

Note that J^- is not complete with respect to the class of basic models which are both sharp and pointed (as model $\mathcal{R}$ for the Russell Example). Indeed, consider formula F,

$$F = \neg(x{:}(P \to Q) \wedge y{:}P \wedge [x{\cdot}y]{:}R)$$

where P, Q, R are distinct propositional letters and x, y justification variables. Obviously, F holds in any basic model $*$ which is sharp and pointed. Imagine a sharp pointed $*$ in which $x{:}(P \to Q)$ and $y{:}P$ hold. In such $*$, $[x{\cdot}y]^* = \{Q\}$, hence both $\neg[x{\cdot}y]{:}R$, and F hold. On the other hand, F is not derivable in J^-, e.g., F fails in the basic model $*$ with $x^* = \{P \to Q\}$, $y^* = \{P\}$, and $t^* = Fm$ for any other t (check closure condition (2)!). So, "sharp and pointed" justification tautologies constitute a proper extension SP of J^-. The problem of finding complete axiomatization of SP was first stated in [5]. This question was answered in [15] along the lines of studying single-conclusion logic of proofs [13,14].

6 Russell Scenario as a *JAM*

Consider the version of J^- in a language with two justification variables w and r, one propositional letter Q, and pointed constant specification CS:

$$c_n{:}A \in CS \quad \text{iff} \quad A \textit{ is an axiom and } n \textit{ is the Gödel number of } A.$$

Define a model $*$ such that

- $Q^* = 1$, i.e., $\models_* Q$;
- $c_n^* = \{A\}$ if A is an axiom and n is the Gödel number $|A|$ of A, and $c_n^* = \emptyset$ otherwise;
- $w^* = r^* = \{Q\}$, e.g., $\models_* r{:}Q$ and $\not\models_* r{:}F$ for any F other than Q (the same for w);
- application is sharp: $[s{\cdot}t]^* = s^* \rhd t^*$.

A *JAM* $\mathcal{R}$ (for Russell's scenario) is $(*, \mathcal{A}, \mathcal{E})$ with

- $\mathcal{A} = \overline{\{w\}}$, i.e., the set of accepted justifications is $\{w\}$, properly closed;
- $\mathcal{E} = \overline{\{r\}}$, i.e., the set of knowledge-producing justifications is $\{r\}$, properly closed.

Though the idea behind $\mathcal{R}$ is quite intuitive, we need to fill in some technical details: extending truth evaluations to all terms and formulas and checking closure conditions.

6.1 Technicalities of the Model

Define $c^*_{|A|} = \{A\}$ for each axiom A of $\mathsf{J}^-(CS)$. Technically, this is an inductive definition with induction on n in c_n.

Base: $n = 0$. Here $c^*_0 = \emptyset$, given 0 is not a Gödel number of any formula.

Inductive step: suppose n is the Gödel number $|F|$ of some formula F. If F is an axiom of J^-, put $c^*_n = \{F\}$. If $F = c_k{:}G$ for some c_k and G, then, by monotonicity of Gödel numbering, $k < n$, hence c^*_k is defined. If $c^*_k = \{G\}$, then $c_k{:}G$ is an axiom of $\mathsf{J}^-(CS)$ and we can define $c^*_n = \{F\}$. In all other cases, $c^*_n = \emptyset$.

Since application is sharp, the evaluation of each term is, at most, a singleton. Together with Boolean truth tables, this determines the truth value of any formula.

Lemma 1. *Each $t \in Tm$ is factive, $\models_* t{:}F \to F$.*

Proof. Induction on t. Assume $\models_* t{:}F$; that means $t^* = \{F\}$. If t is w or r, then F is Q, which is true in the model $*$. If t is a constant, then F is an axiom and hence true in $*$. The induction step corresponds to application, which preserves the truth of justified formulas.

It follows from Lemma 1, that accepted justifications are consistent and knowledge-producing justifications are factive. Therefore, $\mathcal{R} = (*, \mathcal{A}, \mathcal{E})$ is indeed a *JAM*.

Theorem 3. *In model $\mathcal{R}$, sentence Q is true, justified and believed, but not known.*

Proof. In model $\mathcal{R}$, sentence Q is

- true, since $\models_* Q$;
- justified, since $\models_* w{:}Q$;
- believed, since $w \in \mathcal{A}$.

We have to show that Q is not known, i.e., for any justification $g \in \mathcal{A} \cap \mathcal{E}$, $\not\models_* g{:}Q$.

Consider an auxiliary basic model $\bullet$ which is the same as $*$ but with $Q^\bullet = 0$, i.e., the truth value of Q is flipped from 'true' to 'false.' In particular, application in $\bullet$ is sharp.

Lemma 2. *For each justification term t,*

$$t^* = t^\bullet.$$

Proof. The inductive process (based on sharp application) of evaluating all justifications, given evaluations of atomic justifications, operates only with formulas of type $t{:}F$ and starts with the same initial set of such formulas in $*$ and $\bullet$. Hence the results of these processes in $*$ and $\bullet$ coincide.

In particular, for all $g \in \mathcal{A} \cap \mathcal{E}$, $g^* = g^\bullet$, and if $\models_* g{:}Q$, then $\models_\bullet g{:}Q$ as well.

Lemma 3. *Each $g \in \mathcal{A} \cap \mathcal{E}$ is factive in $\bullet$, i.e., $\models_\bullet g{:}F \rightarrow F$.*

Proof. All $g \in \mathcal{A} \cap \mathcal{E}$ are obtained from constants by application. By construction, if $\models_\bullet c{:}X$, then $X \in c^*$ and X is an axiom, hence true. Application obviously preserves factivity.

To complete the proof of Theorem 3, suppose $\mathcal{R} \models g{:}Q$, i.e., $\models_* g{:}Q$, for some $g \in \mathcal{A} \cap \mathcal{E}$. By Lemma 2, $\models_\bullet g{:}Q$, and, by Lemma 1, $\models_\bullet Q$, which is not the case.

6.2 Can Russell's Scenario Be Made Modal?

One could try to express Russell's scenario in a modal language by introducing the justified belief modality

$$\mathbf{J}F \quad \Leftrightarrow \quad \textit{there is } t \in \mathcal{A} \textit{ such that } \models t{:}F,$$

and the knowledge-producing modality

$$\mathbf{E}F \quad \Leftrightarrow \quad \textit{there is } t \in \mathcal{E} \textit{ such that } \models t{:}F,$$

and by stipulating that F is known iff F is both accepted and supported by a knowledge-producing justification:

$$\mathbf{K}F \quad \Leftrightarrow \quad \mathbf{J}F \wedge \mathbf{E}F.$$

This, however, fails, since both $\mathbf{J}Q$ and $\mathbf{E}Q$ hold in $\mathcal{R}$, but $\mathbf{K}Q$ does not. We are facing a Gettier-style phenomenon (cf. [12]), when a proposition is supported by a knowledge-producing justification (hence true), and believed, but not known (since knowledge-producing and accepted justifications for Q are different). This once again illustrates the limitations of modal language in tracking and sorting justifications.

7 Kripke Models and Master Justification

From the Justification Logic point of view, Kripke models may be regarded as a special case of multi-world JAMs[6]; the Kripkean accessibility relation between worlds, uRv, can be recovered by the usual rule *what is believed at u, holds at v*. Moreover, such representation of Kripke models as justification models reveals and formalizes the observation made in [4] that epistemic reading of Kripke models relies on a hidden assumption of (common) knowledge of the model.

The informal argument is as follows. We have to find a justification $m{:}F$ for each knowledge/belief assertion $\Box F$ in a model $\mathcal{K}$. We claim that the model $\mathcal{K}$ itself is such a justification. Indeed, let $u \Vdash \Box F$ in $\mathcal{K}$. Then a complete description

[6] In which we suppress the knowledge-producing component $\mathcal{E}$ to capture beliefs.

of $\mathcal{K}$ yields that at state u, the agent knows/believes F **because the agent knows the model** $\mathcal{K}$ **and knows that** F **holds at all possible worlds**. So, the knowledge/belief-producing evidence for F is delivered by $\mathcal{K}$ itself, assuming the agent is aware of $\mathcal{K}$.

Syntactically, we consider a very basic justification language in which the set of justification terms consists of just one term m, called *master justification*. Think of m as representing a complete description of model $\mathcal{K} = (W, R, \Vdash)$. Specifically, we extend the truth evaluation in $\mathcal{K}$ to justification assertions by stipulating at each $u \in W$

$$\mathcal{K}, u \Vdash m{:}X \quad \textit{iff} \quad \mathcal{K}, v \Vdash X \textit{ for any } v \in R(u) \quad \textit{iff} \quad \mathcal{K}, u \Vdash \Box X.$$

This reading provides a meaningful justification semantics of epistemic assertions in $\mathcal{K}$ via the master justification m representing the whole $\mathcal{K}$. Since a Kripkean agent is logically omniscient, along with $\mathcal{K}$, the agent knows all its logical consequences. Technically, we can assume that the description $\mathcal{K}$ is closed under logical consequence and hence m is idempotent w.r.t. application, $m \cdot m = m$. This condition manifests itself in a special form of the application principle

$$m{:}(A \to B) \to (m{:}A \to m{:}B).$$

On the technical side, a switch from $\Box X$ to $m{:}X$ is a mere transliteration which does not change the epistemic structure of a model. Finally, for each $u \in W$, we define a basic model – maximal consistent set Γ_u in the propositional language with $Tm = \{m\}$:

$$\Gamma_u = \{X \mid u \Vdash X\}.$$

So, from a justification perspective, a Kripke model is a collection of basic models with master justification that represents (common) knowledge of the model.

8 Discussion

Comparisons of justification awareness models with other justification epistemic structures such as Fitting, Mkrtychev, and modular models, can be found in [5]. Technically, basic models and Mkrtychev models may be regarded as special cases of Fitting models. On the other hand, Fitting models can be identified as modular models with additional assumptions, cf. [3]. This provides a natural hierarchy of the aforementioned classes of models:

$$\textit{basic and Mkrtychev models} \subset \textit{Fitting models} \subset \textit{modular models} \subset \textit{JAMs}.$$

Even the smallest class, basic models, is already sufficient for mathematical completeness of justification logics. So, the main idea of progressing to Fitting models, modular models, or JAMs is not a pursuit of completeness but rather a desire to offer natural models for a variety of epistemic situations involving evidence, belief, and knowledge.

*JAM*s do not offer a complete self-contained analysis of knowledge but rather reduce knowledge to knowledge-producing justifications accepted by the agent. This, however, constitutes a meaningful progress; it decomposes knowledge in a way that moves justification objects to the forefront of epistemic modeling. Note that Gettier and Russell examples, clearly indicate which justifications are knowledge-producing or accepted. So *JAM*s fairly model situations in which the corresponding properties of justifications (knowledge-producing, accepted) are given.

There are many natural open questions that indicate possible research directions. Are justification assertions checkable, decidable for an agent? Is the property of a justification to be knowledge-producing checkable by the agent? In multi-agent cases, how much do agents know about each other and about the model? Do agents know each other's accepted and knowledge-producing justifications? What is the complexity of these new justification logics and what are their feasible fragments which make sense for epistemic modeling?

Acknowledgements. The author is grateful to Melvin Fitting, Vladimir Krupski, Elena Nogina, and Tudor Protopopescu for helpful suggestions. Special thanks to Karen Kletter for editing and proofreading this text.

References

1. Artemov, S.: Explicit provability and constructive semantics. Bull. Symbolic Logic **7**(1), 1–36 (2001)
2. Artemov, S.: The logic of justification. Rev. Symbolic Logic **1**(4), 477–513 (2008)
3. Artemov, S.: The ontology of justifications in the logical setting. Stud. Logica. **100**(1–2), 17–30 (2012)
4. Artemov, S.: Knowing the model. Published online at: arXiv:1610.04955 [math.LO] (2016)
5. Artemov, S.: Epistemic modeling with justifications. Published online at: arXiv:1703.07028 [math.LO] (2017)
6. Cresswell, M.J.: Hyperintensional logic. Stud. Logica. **34**(1), 25–38 (1975)
7. Fagin, R., Halpern, J.: Belief, awareness, and limited reasoning. Artif. Intell. **34**(1), 39–76 (1988)
8. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning About Knowledge. MIT Press, Cambridge (1995)
9. Fitting, M.: The logic of proofs, semantically. Ann. Pure Appl. Logic **132**(1), 1–25 (2005)
10. Fitting, M.: Possible world semantics for first-order logic of proofs. Ann. Pure Appl. Logic **165**(1), 225–240 (2014)
11. Fitting, M.: Modal logics, justification logics, and realization. Ann. Pure Appl. Logic **167**(8), 615–648 (2016)
12. Gettier, E.: Is justified true belief knowledge? Analysis **23**, 121–123 (1963)
13. Krupski, V.N.: Operational logic of proofs with functionality condition on proof predicate. In: Adian, S., Nerode, A. (eds.) LFCS 1997. LNCS, vol. 1234, pp. 167–177. Springer, Heidelberg (1997). https://doi.org/10.1007/3-540-63045-7_18
14. Krupski, V.: The single-conclusion proof logic and inference rules specification. Ann. Pure Appl. Logic **113**(1), 181–206 (2002)

15. Krupski, V.: On the sharpness and the single-conclusion property of basic justification models. In: Artemov, S., Nerode, A. (eds.) LFCS 2018. LNCS, vol. 10703, pp. 211–220. Springer, Cham (2018)
16. Kuznets, R., Struder, T.: Justifications, ontology, and conservativity. In: Bolander, T., Braüner, T., Ghilardi, S., Moss, L. (eds.) Advances in Modal Logic, vol. 9, pp. 437–458. College Publications, London (2012)
17. Meyer, J.-J.C., van der Hoek, W.: Epistemic Logic for AI and Computer Science. CUP, Cambridge (1995)
18. Mkrtychev, A.: Models for the logic of proofs. In: Adian, S., Nerode, A. (eds.) LFCS 1997. LNCS, vol. 1234, pp. 266–275. Springer, Heidelberg (1997). https://doi.org/10.1007/3-540-63045-7_27
19. Russell, B.: The Problems of Philosophy. Williams and Norgate, London (1912)
20. Sedlár, I.: Justifications, awareness and epistemic dynamics. In: Artemov, S., Nerode, A. (eds.) LFCS 2013. LNCS, vol. 7734, pp. 307–318. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-35722-0_22
21. Williamson, T.: A note on Gettier cases in epistemic logic. Philos. Stud. **172**(1), 129–140 (2015)

Springer

http://www.springer.com/978-3-319-72055-5

Logical Foundations of Computer Science
International Symposium, LFCS 2018, Deerfield Beach, FL, USA, January 8–11, 2018, Proceedings
Artemov, S.; Nerode, A. (Eds.)
2018, X, 369 p. 17 illus., Softcover
ISBN: 978-3-319-72055-5