A logic of default justifications

Pandzic, Stipe

Published in:
17th International Workshop on Nonmonotonic Reasoning (NMR 2018)

Publication date:
2018

A logic of default justifications

Stipe Pandžić
Department of Theoretical Philosophy, Faculty of Philosophy &
Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Faculty of Science and Engineering
University of Groningen, The Netherlands
s.pandzic@rug.nl

Abstract

We define a logic of default justifications that relies on operational semantics. One of the key features that is absent in standard justification logics is the possibility to weigh different epistemic reasons or pieces of evidence that might conflict with one another. To amend this inadequacy, we develop a semantics for "defeaters": conflicting reasons forming a basis to doubt the original conclusion or to believe an opposite statement. Our logic is able to address interactions of normal defaults without relying on priorities among default rules and introduces the possibility of extension revision for normal default theories.

Introduction

Justification logics provide a formal framework to deal with epistemic reasons. The first justification logic was developed as a logic of arithmetic proofs (LP) by Artemov (2001).[1] Possible world semantics for this logic was first proposed by Fitting (2005a; 2005b) in order to align justification logics within the family of epistemic modal logics. A distinctive feature of justification logic is replacing belief and knowledge modal operators that precede propositions ($\Box P$) by proof terms or, in a generalized epistemic context, justification terms and thereby forming justification assertions $t : P$ that read as "t is a reason that justifies P".

Although justification logic introduced the notions of justification and reason into epistemic logic, it does not formally study the ways of *defeat* among reasons. The importance of defeaters is highlighted by paradigmatic examples from classical literature on defeasible reasoning. The variants of the following example are discussed by Chisholm (1966) and Pollock (1987). Suppose you are standing in a room where you see red objects in front of you. This can lead you to infer that a red-looking table in front of you is in fact red. However, the reason that you have for your conclusion is defeasible. For a typical defeat scenario, suppose you learn that the room you are standing in is illuminated with red light. This gives you a reason to doubt your initial reason to conclude that the table is red, though it would not give you a reason to believe that it is not red. However, if you were to learn, instead, that the table has been painted in white, then you would also have a reason to believe a denial of the claim that the table is red.

The example specifies two different ways in which reasons defeat other reasons: the former is known as *undercut* and the latter as *rebuttal*.[2] Learning additional information

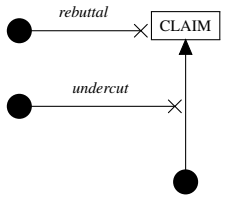

Figure 1: The types of defeat

about the light conditions incurs suspending the applicability of your initial reason to believe that the table is red. In contrast, learning that there is a separate reason to consider that the table is not red will not directly compromise your initial reason itself. The differences between undercutting and rebutting reasons are illustrated in Figure 1.

Only a restricted group of epistemic reasons may be treated as completely immune to defeaters: mathematical proofs. However, they form only a small part of possible reasons to accept a statement and, being a highly-idealized group of reasons, they have rarely been referred to as reasons. Fitting's possible world semantics for justification logics was meant to model not only mathematical and logical truths, but also facts of the world or "inputs from outside the structure" (Fitting 2009, p. 111). Yet the original intent of the first justification logic LP to deal with mathematical proofs, together with the fact that mathematics is cumulative, reflected in its epistemic generalizations. Accordingly, reasons that justify facts of the world were left encapsulated within a framework for non-defeasible mathematical proofs.

Non-mathematical reasons and justifications are commonly held to depend on each other in acquiring their status of "good" reasons and justifications. Still, the questions related to non-ideal reasons have only recently been raised in

[1] The idea of explicit proof terms as a way to find the semantics for provability calculus **S4** dates to 1938 and K. Gödel's lecture published in 1995 (Gödel 1995).

[2] The terminology is originally Pollock's (1987).

the justification logic literature.[3] In the present paper we develop a non-monotonic justification logic with justification terms such that (1) their defeasibility can be tracked from the term structure and (2) other justifications can defeat them by means of an undercut or a rebuttal. Our logic combines techniques from both default logic and justification logic to formalize conflicts of reasons produced in less-than-ideal ways.

Justification logic

The introduction of justifications into modal semantics opened up a possibility to study formal systems for non-defeasible epistemic reasons based on justification logic. These systems include an explicit counterpart to the modal *Truth axiom*: $\Box F \rightarrow F$.[4] Varieties of these systems have been extensively studied and described in e.g. (Kuznets 2000) and (Fitting 2008). Syntactic objects that represent mathematical proofs in LP are more broadly interpreted as epistemic or doxastic reasons by Fitting (2005a; 2005b) and Artemov and Nogina (2005). In order to introduce our system of default reasons, we build upon the existing systems for non-defeasible reasons. In this respect, one can see our strategy as being analogous to the standard default logic approach (Antoniou 1997; Reiter 1980) where agents reason from known or certain information. This section gives preliminaries on one of the logics of non-defeasible reasons.

Since we assume that an agent starts to reason from indefeasible information, we want our underlying logic to represent "factive" or "truth inducing" reasons. However, additional constraints on the system are not necessarily needed to introduce the system of default reasons. Therefore, we do not assume standard axioms and operations that ensure positive or negative introspection. Accordingly, an adequate logical account of factive justifications is the logic $\mathbf{JT}$, a justification logic with the axiom schemes that are explicit analogues of the axiom schemes for modal logic $\mathbf{T}$.[5] After we define the underlying logic, we develop our novel non-monotonic approach to justifications.

Syntax

Syntactically, knowledge operators take the form of justification terms preceding formulas: $t : F$. Given that "t" is a justification term and that "F" is a formula, we write "$t : F$", where t is informally interpreted as a reason or justification for F. We define the set Tm that consists of exactly all justification terms, constructed from variables $x_1, \ldots, x_n, \ldots$ and constants $c_1, \ldots, c_n, \ldots$ by means of operations $\cdot$ and $+$. The grammar of justification terms is given as follows:

$$t ::= x \mid c \mid (t_1 \cdot t_2) \mid (t_1 + t_2)$$

where x is a variable denoting an unspecified justification and c is a proof constant. Proof constant c is atomic within the system. For a justification term t, a set of subterms $Sub(t)$ is defined by induction on the construction of t. Formulas of $\mathbf{JT}$ are defined by the following grammar:

$$F ::= \top \mid P \mid (F_1 \rightarrow F_2) \mid (F_1 \vee F_2) \mid (F_1 \wedge F_2) \mid \neg F \mid t : F$$

where $P \in \mathcal{P}$ and $\mathcal{P}$ is an enumerable set of atomic propositional formulas and $t \in Tm$. The set Fm consists of exactly all formulas.

Axioms and rules of JT

We can now define the logic of non-defeasible reasons $\mathbf{JT}$. The logic $\mathbf{JT}$ is the weakest logic with "truth inducing" justifications containing axiom schemes for two basic operations $\cdot$ and $+$.[6] These are the axioms and rules of $\mathbf{JT}$:

A0 *All the instances of propositional logic tautologies from* Fm

A1 $t : (F \rightarrow G) \rightarrow (u : F \rightarrow (t \cdot u) : G)$ (Application)

A2 $t : F \rightarrow (t + u) : F$; $\quad u : F \rightarrow (t + u) : F$ (Sum)

A3 $t : F \rightarrow F$ (Factivity)

R0 *From* F *and* $F \rightarrow G$ *infer* G (Modus ponens)

R1 *If* F *is an axiom instance of A0-A3 and* $c_n, c_{n-1} \ldots, c_1$ *proof constants, then infer* $c_n : c_{n-1} \ldots c_1 : F$ (Iterated axiom necessitation)

Proof constants are justifications of basic logic truths. In justification logics, basic truths are taken to be justified (at any depth) by virtue of their status within a system and their justifications are not further analyzed. A set of instances of such canonical formulas in justification logic is called *Constant Specification* ($\mathcal{CS}$) set.

Definition 1 (Constant specification)**.** *The Constant Specification set is the set of instances of rule R1.*

$$\mathcal{CS} = \{c_n : c_{n-1} \ldots c_1 : A \mid A \text{ is an axiom instance of A0-A3, } c_n, c_{n-1}, \ldots, c_1 \text{ are proof constants and } n \in \mathbb{N}\}$$

The use of constants in R1 above is unrestricted. In such format, the rule generates a set of formulas where each axiom is justified by any constant at any depth. The set of formulas obtained in this way is called *Total Constant Specification* ($\mathcal{TCS}$). A more appropriate name for the logic above would therefore be $\mathbf{JT_{TCS}}$. It is possible to put restrictions on the use of constants in rule R1 in order to consider a limited class of $\mathcal{CS}$-sets. We restrict the constant specification

[3] The first proposed formalism that includes the idea of evidence elimination specific to a multi-agent setting is by Renne (2012). Baltag, Renne and Smets (2012; 2014) bring together ideas from belief revision and dynamic epistemic logic and offer an account of good and conclusive evidence. Several approaches ((Milnikel 2014), (Kokkinis et al. 2015), (Kokkinis, Ognjanović, and Studer 2016) and (Ognjanović, Savić, and Studer 2017)) start from the idea of merging probabilistic degrees of belief with justification logic, while (Fan and Liau 2015) and (Su, Fan, and Liau 2017) develop a possibilistic justification logic.

[4] In fact, in (Fitting 2008, p. 156) we find three different truth axiom schemes.

[5] Justification logic $\mathbf{JT}$ was first introduced by Brezhnev (2001). Justification logics with equivalent axiom schemes to the logic we define in this section are also defined and investigated in (Kuznets 2000) and (Fitting 2008). In future work, adding the axioms of positive and negative introspection could be considered.

[6] As Fitting (2005b; 2008) shows, we can also technically consider dropping the operator $+$ from our language. In this way we obtain the logic that he calls LP$^-$(T) (Fitting 2008, p. 162).

set $\mathcal{CS}$ following a simple intuition that each axiom instance has its own proof constant.[7]

Restriction 2. *$\mathcal{CS}$ is*

- *Axiomatically appropriate: for each axiom instance A, there is a constant c such that $c : A \in \mathcal{CS}$ and for each formula $c_n : c_{n-1} \dots c_1 : A \in \mathcal{CS}$, such that $n \geq 1$, $c_{n+1} : c_n : c_{n-1} \dots c_1 : A \in \mathcal{CS}$ for some c_{n+1};*
- *Injective: Each proof constant c justifies at most one formula.*

The logic $\mathbf{JT_{CS}}$ is defined by replacing the iterated axom necessitation rule of $\mathbf{JT_{TCS}}$ with the following rule dependent on Restriction 2:

R1* *If F is an axiom instance of A0-A3 and $c_n, c_{n-1} \dots, c_1$ proof constants such that $c_n : c_{n-1} : \dots c_1 : F \in \mathcal{CS}$, then infer $c_n : c_{n-1} : \dots c_1 : F$*

We say that the formula F is $\mathbf{JT_{CS}}$-provable ($\mathbf{JT_{CS}} \vdash F$) if F can be derived using the axioms A0-A3 and rules R0 and R1*.

Semantics

The semantics for $\mathbf{JT_{CS}}$ is an adapted version of the semantics for the logic of proofs ($\mathbf{LP}$) given by Mkrtychev (1997).[8]

Definition 3 ($\mathbf{JT_{CS}}$ model)**.** *We define a function* reason assignment *based on $\mathcal{CS}$ $*(\cdot) : Tm \to 2^{Fm}$, a function mapping each term to a set of formulas from Fm. It satisfies the following conditions:*

1. *If $F \to G \in *(t)$ and $F \in *(u)$, then $G \in *(t \cdot u)$*
2. *$*(t) \cup *(u) \subseteq *(t + u)$*
3. *If $c : F \in \mathcal{CS}$, then $F \in *(c)$*

A truth assignment $v : \mathcal{P} \to \{True, False\}$ is a function assigning truth values to propositional formulas in $\mathcal{P}$. We define the interpretation $\mathcal{I}$ as a pair $(v, *)$. For an interpretation $\mathcal{I}$, $\models$ is a truth relation on the set of formulas of $\mathbf{JT_{CS}}$.

For any formula $F \in Fm$, $\mathcal{I} \models F$ iff

- For any $P \in \mathcal{P}$, $\mathcal{I} \models P$ iff $v(P) = True$
- $\mathcal{I} \models \neg F$ iff $\mathcal{I} \not\models F$
- $\mathcal{I} \models F \to G$ iff $\mathcal{I} \not\models F$ or $\mathcal{I} \models G$
- $\mathcal{I} \models F \vee G$ iff $\mathcal{I} \models F$ or $\mathcal{I} \models G$
- $\mathcal{I} \models F \wedge G$ iff $\mathcal{I} \models F$ and $\mathcal{I} \models G$
- $\mathcal{I} \models t : F$ iff $F \in *(t)$

The interpretation $\mathcal{I}$ is *reflexive*, which means that the truth relation for $\mathcal{I}$ fulfills the following condition:

- For any term t and any formula F, if $F \in *(t)$, then $\mathcal{I} \models F$.

Definition 4 ($\mathbf{JT_{CS}}$ consequence relation)**.** *$\Sigma \models F$ iff for all reflexive interpretations $\mathcal{I}$, if $\mathcal{I} \models B$ for all $B \in \Sigma$, then $\mathcal{I} \models F$.*

Due to Restriction 2, the consequence relation for $\mathbf{JT_{CS}}$ is weaker than the $\mathbf{JT_{TCS}}$ consequence relation.

Definition 5 ($\mathbf{JT_{CS}}$ closure)**.** *JT_{CS} closure is given by $Th^{JT_{CS}}(\Gamma) = \{F | \Gamma \models F\}$, for a set of formulas $\Gamma \subseteq Fm$ and the JT_{CS} consequence relation $\models$ defined above.*

For any closure $Th^{JT_{CS}}(\Gamma)$, it follows that $\mathcal{CS} \subseteq Th^{JT_{CS}}(\Gamma)$.

We can prove that the compactness theorem holds for the $\mathbf{JT_{CS}}$ semantics.[9] Compactness turns out to be a useful result in defining the operational semantics of default reason terms. We first say that a set of formulas Γ is $\mathbf{JT_{CS}}$ *satisfiable* if there is an interpretation $\mathcal{I}$ that meets $\mathcal{CS}$ (via the third condition of Def. 3) for which all the members of Γ are true. A set Γ is $\mathbf{JT_{CS}}$-finitely satisfiable if every finite subset Γ' of Γ is $\mathbf{JT_{CS}}$ satisfiable.

Theorem 6 (Compactness)**.** *A set of formulas is $\mathbf{JT_{CS}}$ satisfiable iff it is $\mathbf{JT_{CS}}$-finitely satisfiable.*

Proof. See the Appendix. □

A logic of default justifications

In this section, we develop a system based on $\mathbf{JT_{CS}}$ in which agents form default justifications reasoning from an incomplete knowledge base. Justification logic $\mathbf{JT_{CS}}$ is capable of representing the construction of a new piece of evidence out of existing ones by application ("$\cdot$") or sum ("+") operation. However, to extend an incomplete $\mathbf{JT_{CS}}$ theory, we need to import reasons that are defeasible. We come up with both a way in which such reasons are imported and a way in which they might get defeated by introducing concepts familiar from defeasible reasoning literature into justification logic.

We start from the above-defined language of the logic $\mathbf{JT_{CS}}$ and develop a new variant of justification logic $\mathbf{JT_{CS}}$ that enables us to formalize the import of reasons outside the structure as well as to formalize *defeaters* or reasons that question the plausibility of other reasons.

Our logical framework of defeasible reasons represents both factive reasons produced via the axioms and rules of $\mathbf{JT_{CS}}$ and plausible reasons based on default assumptions

[7] For example, one such constant specification is defined by Artemov (2018, p. 31): "$c_n : A \in \mathcal{CS}$ iff A is an axiom and n is the Gödel number of A". The choice of $\mathcal{CS}$ is not trivial. If we define an empty $\mathcal{CS}$, that is, $\mathbf{JT}_\emptyset$, we eliminate logical awareness for agents, while defining an infinite $\mathcal{CS}$ imposes logical omniscience. To ensure that standard properties as *Internalization* (Artemov 2001) hold, $\mathcal{CS}$ has to be axiomatically appropriate. Moreover, different restrictions could affect complexity results, as discussed in e.g. (Milnikel 2007).

[8] The condition for justifications of the type '$!t$' are not needed in the $\mathbf{JT_{CS}}$ semantics. Mkrtychev's model can be thought of as a single world justification model. Since the notion of defeasibility introduced in the next section turns on the incompleteness of available reasons, our system eliminates worries about the trivialization of justification assertions that otherwise arise from considering justifications as modalities in a single-world model.

[9] A compactness proof for $\mathbf{LP}$ satisfiability in possible world semantics is given in (Fitting 2005b).

that "usually" or "typically" hold for a restricted context.[10] We follow the standard way (Reiter 1980) of formalizing default reasoning through default theories to extend the logic of factive reasons with defeasible reasons. Building on the syntax of $\mathbf{JT_{CS}}$, we introduce the definition of the *default theory*:

Definition 7 (Default Theory)**.** *A default theory T is defined as a pair (W, D), where the set W is a finite set of $\mathbf{JT_{CS}}$ formulas and D is a countable set of default rules.*

Each default rule is of the following form:

$$\delta = \frac{t : F :: (u \cdot t) : G}{(u \cdot t) : G}.$$

The informal reading of the default δ is: "If t is a reason for F, and it is consistent to assume that $(u \cdot t)$ is a reason for G, then $(u \cdot t)$ is a defeasible reason to believe G". The formula $t : F$ is called the *prerequisite* and $(u \cdot t) : G$ is both the *consistency requirement*[11] and the *consequent* of the default rule δ. We refer to each of the respective formulas as $pre(\delta)$, $req(\delta)$ and $cons(\delta)$. For the set of all consequents from the entire set of defaults D, we use $cons(D)$. The default rule δ introduces a unique reason term u, which means that, for a default theory T, the following holds:

1. For any formula $t : F \in Th^{JT_{CS}}(W)$, $u \neq t$ and
2. For any other default rule $\delta' \in D$ such that $\delta' = \frac{t':F'::(u'\cdot t'):G'}{(u'\cdot t'):G'}$, if $F \neq F'$ or $G \neq G'$, then $u \neq u'$.[12]

The reason term u witnesses the defeasiblity of the *prima facie* reason $(u \cdot t)$ for G. Whether a reason actually becomes defeated or not depends on other default-reason formulas from $cons(D)$. Other defaults might question both the plausibility of the default reason u and the plausibility of the proposition G.

A formal way of looking at a default reason of this kind is that $(u \cdot t)$ codifies the default step we apply on the basis of the known reason t. A distinctive feature of such rules is generating justification terms as if it were the case that $cons(\delta)$ was inferred by using an instance of the application axiom: $u : (F \to G) \to (t : F \to (u \cdot t) : G)$. The difference is that an agent cannot ascertain that an available reason justifies applying the conditional $F \to G$ without restrictions. Still, sometimes a conclusion must be drawn without being able to remove all of the uncertainty as to whether the relevant conditional actually applies or not. In such cases, an agent turns to a plausible assumption of a justified "defeasible" conditional $F \to G$ that holds only in the absence of any information to the contrary. While the internal structure of the default reason $(u \cdot t)$ indicates that it is formed on the basis of the formula $u : (F \to G)$, the defeasibility of $(u \cdot t)$ lies in the fact that the formula $u : (F \to G)$ is not a part of the knowledge base.

One can think of our use of the operation "$\cdot$" in default rules as the same operation that is used in the axiom A1, only being applied on an incomplete $\mathbf{JT_{CS}}$ theory. Similarly, we can follow Reiter (1980, p. 82) and Antoniou (1997, p. 21) in thinking of a standard default rule such as $\frac{A:B}{B}$ as merely saying that an implication $A \wedge \neg C \wedge \neg D \cdots \to B$ holds, provided that we can establish that a number of exceptions $C, D, \ldots$ does not hold. However, if the rule application context is defined sufficiently narrowly, the rule is classically represented as an implication $A \to B$. Generalizing on such interpretation of defeasibility, our defaults with justification assertions can be represented as instantiations of the axiom A1 applied in a sufficiently narrow application context.

Analogous to standard default theories, we take the set of facts W to be underspecified with respect to a number of facts that would otherwise be specified for a complete $\mathbf{JT_{CS}}$ interpretation. Besides simple facts, our underlying logic contains justification assertions. To deal with justification assertions, a complete $\mathbf{JT_{CS}}$ interpretation would also further specify whether a reason is acceptable as a justification for some formula. Therefore, except the usual incomplete specification of known propositions, default justification theories are also incomplete with respect to the actual specification of the reason assignment function. For our default theory, this means that, except the valuation v, default rules need to approximate an actual reason-assignment function $*(\cdot)$.

In "guessing" what a true model is, every default rule introduces a reason term whose structure codifies an application operation step from an unknown justified conditional. For example, in rule δ above, we rely on the justified conditional $u : (F \to G)$. Even though this justified conditional is not a part of the rule δ itself, it is the underlying assumption on the basis of which we are able to extend an incomplete knowlede base. Each underlying assumption of this kind can be made explicit by means of a function *default conditional assignment*: $\#(\cdot) : D \to Fm$. The function maps each default rule to a specific justified conditional as follows:

$$\#(\delta_i) = u_n : (F \to G),$$

where $\delta_i \in D$ and $\delta_i = \frac{t_k:F::(u_n\cdot t_k):G}{(u_n\cdot t_k):G}$, for some reason terms t_k and u_n and some formulas F and G.

A set of all such underlying assumptions of default rules is called *Default Specification* ($\mathcal{DS}$) set.

Definition 8 (Default specification)**.** *For a default theory $T = (W, D)$, justified defeasible conditionals are given by the Default Specification set:*

$$\mathcal{DS} = \#[D] = \{u_n : (F \to G) \mid \#(\delta_i) = u_n : (F \to G) \text{ and } \delta_i \in D\}.$$

The use of underlying assumptions from $\mathcal{DS}$ is responsible for the non-monotonic character of default reasons and contrasts our default rules with the standard application operation represented by the axiom A1. The extended meaning of

[10] For a logical account of typicality based on ranked models and preferential reasoning, see Propositional Typicality Logic (PTL) developed by Booth, Meyer, and Varzinczak (2012). In PTL, a typicality operator is added to propositional logic and interpreted in terms of ranked models to formally capture the most typical situations in which a given formula holds.

[11] In order to avoid any misunderstanding, we avoid the name *justification* for the formula $req(\delta)$ since justification logic terms are commonly known as justifications.

[12] Similarly, Artemov (2018, p. 30) introduces "single-conclusion" (or "pointed") justifications to enable handling "justifications as objects rather than as justification assertions".

the application operation via default rules will be referred to as **default application**. Extending the interpretation of the application operation "$\cdot$" can be formally captured by the following definition:

Definition 9 (Default Application)**.** *For a default rule $\delta \in D$, if $u : (F \to G) = \#(\delta)$ and if $t : F = pre(\delta)$, then $(u \cdot t) : G = cons(\delta)$.*

Let us again consider the red-looking-table example from the Introduction to see how *prima facie* reasons and their defeaters are imported through default rules.

Example 10. *Let R be the proposition "the table is red-looking" and let T be the proposition "the table is red". Take t_a and u_a to be some specific individual justifications. The reasoning whereby one accepts the default reason $(u_a \cdot t_a)$ might be described by the following default rule:*

$$\delta_a = \frac{t_a : R :: (u_a \cdot t_a) : T}{(u_a \cdot t_a) : T}.$$

We can informally read the default as follows: "If you have a reason to believe that a table is red looking and it is consistent for you to assume that this gives you a reason supporting the claim that the table is red, then you have a defeasible reason to conclude that the table is red". Suppose you then get to a belief that "the room you are standing in is illuminated with red light", a proposition denoted by L. For some specific justifications t_b and u_b, the following rule gives you an undercutting reason for $(u_a \cdot t_a)$:

$$\delta_b = \frac{t_b : L :: (u_b \cdot t_b) : \neg[u_a : (R \to T)]}{(u_b \cdot t_b) : \neg[u_a : (R \to T)]},$$

where the rule is read as "If you have a reason to believe that the lighting is red and it is consistent for you to assume that this gives you a reason to deny your reason to conclude that the red-looking table is red, then you have a defeasible reason that denies your reason to conclude that the red-looking table is red". The formula $cons(\delta_b)$ denies the basis for the inference that led you to conclude $cons(\delta_a)$, although note that it is not directly inconsistent with it. In the next subsection we define what undercutting defeaters are semantically.

Suppose that instead of learning about the light conditions in the room as in δ_b, you learn that the table has been painted white. This would also prompt a rebutting defeater - a separate reason to believe the contradicting proposition $\neg T$. Let W denote the proposition "the table is painted white" and let t_c and u_c be some specific justifications. We have the following rule:

$$\delta_c = \frac{t_c : W :: (u_c \cdot t_c) : \neg T}{(u_c \cdot t_c) : \neg T}.$$

The rule reads as "If you have a reason to believe that the table has been painted white and it is consistent for you to assume that this gives you a reason supporting the claim that the table is not red, then you have a defeasible reason to conclude that the table is not red". Note that the formula $cons(\delta_c)$ does not directly mention any of the subterms of $(u_a \cdot t_a)$. The defeat among the reasons $(u_a \cdot t_a)$ and $(u_c \cdot t_c)$ comes from the fact that they cannot together consistently extend an incomplete $\mathbf{JT_{CS}}$ theory.

The entire example can be described by the following default theory $T_0 = (W_0, D_0)$, where $W_0 = \{t_a : R, t_b : L, t_c : W\}$ and $D_0 = \{\delta_a, \delta_b, \delta_c\}$.

Each defeater above is itself defeasible and considered to be a *prima facie* reason. The way in which *prima facie* reasons interact is further specified through their role in the operational semantics.

Operational semantics of default justifications

Between the two types of defeaters, the semantics of rebutting justifications is more straightforward since it rests on the known mechanism of multiple extensions used in standard default theories. What requires additional explanation is the semantics of undercutting defeaters. Notice that each formula $\#(\delta)$ has the format of a justified material conditional. This formula is not a part of a default inference δ itself, but the default application described by δ depends on assuming a reason for that conditional and the justification assertion $cons(\delta)$ encodes this assumption in the internal structure of the resulting reason term. This brings to attention the following possibility: a knowledge base may at the same time contain justified formulas of the type $t : F$, $(u \cdot t) : G$ and $v : \neg[u : (F \to G)]$, without the knowledge base being inconsistent. Although the application axiom A1 does not say that $t : F$ and $(u \cdot t) : G$ together entail the formula $u : (F \to G)$, the occurrence of the formulas $t : F$, $(u \cdot t) : G$ and $v : \neg[u : (F \to G)]$ together is not significant in standard justification logic. It only becomes significant with default application.[13]

The extension of the application operation to its defeasible variant opens new possibilities for a semantics of justifications. In particular, it enables reasoning that is not regimented by the standard axioms A1 and A2 of basic justification logic (Artemov 2008, p. 482). For instance, if a set of $\mathbf{JT_{CS}}$ formulas contains both a *prima facie* reason t and its defeater u, then the set containing a conflict of justifications does not support concatenation of reasons by which $t : F \to (t+u) : F$ holds for any two terms t and u. In other words, the possibility of a conflict between reasons eliminates the monotonicity property of justifications assumed in the sum axioms (A2).

The logic of default justifications we develop here relies on the idea of operational semantics for standard default logics presented in (Antoniou 1997). Here is an informal description of the key operational semantics steps. First, default reasons are taken into consideration at face value. After the default reasons have been taken together, we check dependencies among them in order to find out what are the non-defeated reasons. Finally, a rational agent includes in its knowledge base only acceptable pieces of information that are based on those reasons that are ultimately non-defeated.

The basis of operational semantics for a default theory $T = (W, D)$ is the procedure of collecting new, reason-

[13] Notice that a ($\mathbf{JT_{CS}}$-closed) knowledge base that contains the formulas $t : F$ and $(u \cdot t) : G$, also contains the formula $((c \cdot t) \cdot (u \cdot t)) : (F \to G)$, assuming that the constant c justifies the axiom $F \to (G \to (F \to G))$. This is so regardless of whether $u : (F \to G)$ is also in the knowledge base or not.

based information from the available defaults. A *sequence* of default rules $\Pi = (\delta_0, \delta_1, \ldots)$ is a possible order in which a list of default rules without multiple occurrences from D is applied (Π is possibly empty). Applicability of defaults is determined in the following way: for a set of $\mathbf{JT_{CS}}$-closed formulas Γ we say that a default rule $\delta = \frac{t:F::(u\cdot t):G}{(u\cdot t):G}$ is applicable to Γ iff

- $t : F \in \Gamma$ and
- $\neg(u \cdot t) : G \notin \Gamma$.

Reasons are brought together in the set of $\mathbf{JT_{CS}}$ formulas that represents the current evidence base:

Definition 11. $In(\Pi) = Th^{JT_{CS}}(W \cup \{cons(\delta) \mid \delta \textit{ occurs in } \Pi\})$.

The set $In(\Pi)$ collects reason-based information that is yet to be determined as acceptable or unacceptable depending on the acceptability of reasons and counter-reasons for formulas.

We need to further specify sequences of defaults that are significant for a default theory T: default processes. For a sequence Π, the initial segment of the sequence is denoted as $\Pi[k]$, where k stands for the number of elements contained in that segment of the sequence and where k is a minimal number of defaults for the sequence Π. Any segment $\Pi[k]$ is also a sequence. Intuitively, the set of formulas $In(\Pi)$ represents an updated incomplete knowledge base W where the new information is not yet taken to be granted. Using the notions defined above, we can now get clear on what a default process is:

Definition 12 (Process)**.** *A sequence of default rules* Π *is a* process *of a default theory* $T = (W, D)$ *iff every* k *such that* $\delta_k \in \Pi$ *is* applicable *to the set* $In(\Pi[k])$*, where* $\Pi[k] = (\delta_0, \ldots \delta_{k-1})$.

We will use default specification sets that are relativized to default processes:

$$\mathcal{DS}^{\Pi} = \{u_n : (F \to G) \mid \#(\delta_i) = u_n : (F \to G) \text{ and } \delta_i \in \Pi\}.$$

The kind of process that we are focusing on here is called *closed* process and we say that a process Π is closed iff every $\delta \in D$ that is applicable to $In(\Pi)$ is already in Π. For default theories with a finite number of defaults, closure for any process Π is obviously guaranteed by the applicability conditions. However, if a set of defaults is infinite, then this is less-obvious.

Lemma 13 (Infinite Closed Process)**.** *For a theory* $T = (W, D)$ *and infinitely many* k*'s, an infinite process* Π *is closed iff for every default rule* δ_k applicable *to the set* $In(\Pi[k])$*,* $\delta_k \in \Pi$.

Proof. From the compactness of $\mathbf{JT_{CS}}$ semantics we have that if a set $In(\Pi[k]) \cup \{req(\delta)\}$ is satisfiable for all the finite k's, it is also satisfiable for infinitely many k's. Therefore the applicability conditions for a rule δ are equivalent to the finite case. □

Besides the standard process of collecting new information, we need to explain the way in which an agent decides on the acceptability of reasons. We have already introduced the extended meaning of the application operation for a default theory T. Now we show how default application is essential to the operational semantics of default reasons. Ideally, an agent has all the factive reasons valid under some interpretation $\mathcal{I}$. In contrast, in reasoning from an incomplete knowledge base W, a closure $Th^{JT_{CS}}(W)$ is typically underspecified as to whether a reason t is acceptable for a formula F. In such context, reasoning starts from defeasible justification assertions in $\mathcal{DS}$ as the only available resource to approximate a reason assignment function that actually holds.

Notice that $\mathcal{DS}$ can be an inconsistent set of $\mathbf{JT_{CS}}$ formulas and that an agent needs to find out which reasons prevail in a conflicting set of reasons. One way in which reasons may conflict with each other is captured by the definition of undercut:

Definition 14 (Undercut)**.** *A reason* u *undercuts reason* t *being a reason for a formula* F *in a set of* $\mathbf{JT_{CS}}$*-closed formulas* $\Gamma \subseteq In(\Pi[k])$ *iff* $\bigvee_{(v)\in Sub(t)} u : \neg[v : (G \to H)] \in \Gamma$ *and* $v : (G \to H) \in \mathcal{DS}^{\Pi}$.

For a set Γ such that $Th^{JT_{CS}}(\Gamma)$ contains some reason u that undercuts t we say that Γ undercuts t. We can think of Γ as a set of reasons against which we test the reason t being reason for the formula F. This is further elaborated in the semantics of acceptability of reasons. We now define conflict-free sets of formulas:[14]

Definition 15 (Conflict-free sets)**.** *A set of* $\mathbf{JT_{CS}}$*-closed formulas* Γ *is conflict-free iff* Γ *does not contain both a formula* $t : F$ *with an undercut reason* t *and its undercutter* $u : G$.

As stated before, the set W contains certain information and this means that any information from W is always acceptable regardless of what has been collected later on. Therefore, any set of formulas Γ that extends the initial information contains W. To decide whether a consequent of a default δ is acceptable, an agent looks at those sets of reasons that can be defended against all the available counter-reasons. According to that, an agent looks at finding a defensible set of justified formulas among all certain information taken together with the consequents of the applicable defaults rules. Therefore, for a default theory $T = (W, D)$, an agent always considers potential extension sets of $\mathbf{JT_{CS}}$ formulas that meet the following conditions:

1. $W \subseteq \Gamma$ and
2. $\Gamma \subseteq \{W \cup cons(\delta) \mid \delta \text{ occurs in } \Pi_i\}$,

where Π_i is a closed process of T. For any potentially acceptable set Γ we define the notion of acceptability of a justified formula $t : F$:

[14] In characterizing sets of $\mathbf{JT_{CS}}$ formulas we use the terminology of Dung's (1995) abstract argumentation frameworks whenever possible. Abstract argumentation frameworks treat conflicts between arguments and they naturally overlap with our idea of conflicting reasons in many ways.

Definition 16 (Acceptability)**.** *For a default theory $T = (W, D)$, a formula $t : F \in cons(\Pi)$ is acceptable w.r.t. a set of $\mathbf{JT_{CS}}$ formulas Γ iff for each undercutting reason u for t being a reason for F such that $u : G \in In(\Pi)$, $Th^{JT_{CS}}(\Gamma)$ undercuts u being a reason for G.*

Informally, an agent has yet to test any potential extension against all the other available reasons before it can be considered as an admissible extension of the knowledge base.

Definition 17 (Admissible Extension)**.** *A potential extension set of $\mathbf{JT_{CS}}$ formulas Γ is an admissible extension of a default theory $T = (W, D)$ iff $Th^{JT_{CS}}(\Gamma)$ is conflict-free and if each formula $t : F \in \Gamma$ is acceptable w.r.t. Γ.*

After considering all the available reasons, an agent accepts only those defeasible statements that can be defended against all the available reasons against these statements.

The two latter definitions introduce the idea of "external stability" of knowledge bases (Dung 1995, p. 323) into default logic by taking into account all the reasons that question the plausibility of other reasons. In addition to that, our operational semantics prompts an implicit revision procedure. Any new default rule that is applicable to the set of formulas $In(\Pi[k])$ potentially makes changes to what an agent considered to be acceptable relying on the set of formulas $In(\Pi[k-1])$. Before we show this on the formalized example from the beginning of this section, we introduce the idea of default extension for a default theory T. Extension is the fundamental concept in defining logical consequence in standard default theories. We think of preferred extensions as maximal plausible world views based on the acceptability of reasons:

Definition 18 (Preferred Extension)**.** *For a default theory $T = (W, D)$, an admissible extension set of $\mathbf{JT_{CS}}$ formulas Γ, $Th^{JT_{CS}}(\Gamma)$ is a preferred extension of a default theory T iff for any other admissible extension Γ', $\Gamma \not\subset \Gamma'$.*

In other words, preferred extensions are maximal admissible extensions with respect to set inclusion. The existence of preferred extensions is universally defined for default theories. To ensure that this result also holds for the case of an infinite number of default rules and infinite closed processes, we make use of Zorn's lemma and restate it as follows:

Lemma 19 (Zorn)**.** *For every partially ordered set A, if every chain of (totally ordered subset of) B has an upper bound, then A has a maximal element.*

Theorem 20 (Existence of Preferred Extension)**.** *Every default theory $T = (W, D)$ has at least one preferred extension.*

Proof. If W is inconsistent, then for any default δ, negation of the consistency requirement $req(\delta)$ is contained in $Th^{JT_{CS}}(W)$ and the only closed process Π is the empty sequence. Therefore, the only potential and admissible extension is W itself and T has a unique preferred extension $Th^{JT_{CS}}(W)$ containing all the formulas of $\mathbf{JT_{CS}}$.

Assume that W is consistent. In general, if there is a finite number of default rules in D, any closed process Π of T is also finite. Admissible extensions obtained from closed processes form a complete partial order with respect to $\subseteq$. Since there are only finitely many admissible sets, any admissible set Γ has a maximum Γ' within a totally ordered subset of a set of all admissible sets. Therefore, $\Gamma \subseteq \Gamma'$ and $Th^{JT_{CS}}(\Gamma')$ is a preferred extension of T.

For the case where D is infinite and closed processes $\Pi_1, \Pi_2, \ldots$ are infinite, there is again a complete partial order formed from a set of all admissible sets. The argument for finite processes does not account for the case where Γ', the union of admissible sets $\Gamma_1, \Gamma_2, \ldots$, could be contained in some Γ'' for an ever increasing sequence $\Gamma_1, \Gamma_2, \ldots$. We first state that Γ', the union of an ever increasing sequence of admissible sets $\Gamma_1, \Gamma_2, \ldots$, is also an admissible set. To ensure this, we turn to its subsets. That is, if Γ' was not admissible, then some of its subsets Γ_n for $n \geq 1$ would not be conflict-free or would contain a formula that is not acceptable, but this contradicts the assumption that Γ_n is admissible. Now, for the set of all admissible sets ordered by $\subseteq$, any chain (totally ordered subset) has an upper bound, that is, the union of its members $\Gamma' = \bigcup_{n=1}^{\infty} \Gamma_n$. According to Lemma 19, there exists a maximal element and, therefore a preferred extension of T.

□

The semantics of defeasible reasons enables us to define additional types of extensions that are not necessarily based on the admissibility of reasons. One of them is stable extension familiar from formal argumentation theory:

Definition 21 (Stable Extension)**.** *For a default theory $T = (W, D)$ and its closed processes Π and Π', a stable extension is a $\mathbf{JT_{CS}}$ closure of a potential extension $\Gamma \subset In(\Pi)$ such that* **(1)** *$Th^{JT_{CS}}(\Gamma)$ undercuts all the formulas $t : F \in In(\Pi)$ outside $Th^{JT_{CS}}(\Gamma)$ and* **(2)** *for any formula $u : G \in \Gamma'$ such that $\Gamma' \subset In(\Pi')$ and $u : G \notin In(\Pi)$, it holds that $\Gamma \cup \{u : G\}$ is $\mathbf{JT_{CS}}$ inconsistent.*

The intuition behind the definition is that every reason left outside the accepted set of reasons is attacked. For our logic, this means that for every justification assertion outside of an extension, the extension undercuts one of its subterms and/or it contains a justification assertion inconsistent with it. We can check that in the red-looking-table example, stable and prefer extension coincide. Formally, theory T_0 has a unique stable and preferred extension $Th^{JT_{CS}}(W_0 \cup \{cons(\delta_b), cons(\delta_c)\})$. Moreover, note that the process (δ_a, δ_b) includes a revision of its respective admissible extension.

Stable extensions are not universally defined for any default theory T. Consider the following theory $T_1 = (W_1, D_1)$, where $W_1 = \{t : F\}$ and D_1 contains the default rules

$$\delta_1 = \frac{t : F :: (u \cdot t) : G}{(u \cdot t) : G} \text{ and}$$

$$\delta_2 = \frac{(u \cdot t) : G :: (v \cdot (u \cdot t)) : \neg[u : (F \rightarrow G)]}{(v \cdot (u \cdot t)) : \neg[u : (F \rightarrow G)]}.$$

While T_1 has a preferred extension $Th^{JT_{CS}}(W)$, it has no stable extension. This result conforms to similar results about preferred and stable semantics in abstract argumentation frameworks. In fact, T_1 is a justification logic formal-

ization of the concept of self-defeat, which is notorious in argumentation framworks.

In addition, we can easily add other significant notions of extensions, analogous to those in (Dung 1995). In particular, we can define variants of Dung's (1995, p. 329) *complete* and *grounded* extension. Different extensions definitions will enable us to give different corresponding characterizations of logical consequence. This will lead to proofs of additional theorems and fully establish the role of justification logic within the study of non-monotonic reasoning.

Related and future work

The above suggested connections between default justification logic and abstract argumentation frameworks are currently being investigated. Standard justification logics are known for their connection to modal logics. Artemov (2001) provided a proof of the *Realization Theorem* that connects the logic of arithmetic proofs LP with the modal logic S4. The result has been followed up by similar theorems for many other modal logics with known "explicit" justification counterparts.[15] As it stands now, default justification logic can be considered to provide explicit justification logic counterparts to (a subclass of) abstract argumentation frameworks. A proof of this conjecture is a part of the future work.

Further developments are possible starting from the basic logic of default justifications. On the technical side of our logic, we used only the expressiveness of normal default rules and we still need to investigate how to add non-normal default rules. In the general context of default logics, our logic introduces some new technical properties for normal default theories that are still to be thoroughly described. Among them are revision of extensions and interaction of different defaults that does not rely on their preference orderings, as commonly done in default logic (Delgrande and Schaub 2000). An extensive account of default reasons that makes use of preference orderings on defaults is developed by Horty (2012). Horty's logic is based on a propositional language and develops from a different notion of reasons, which makes it incomparable to our logic where reasons are explicitly featured in the language itself.

Our work provides a complementary addition to the study of less-than-ideal reasons in justification logic. Among related approaches, the logic of conditional probabilities developed by Ognjanović, Savić, and Studer (2017) introduces a way to model non-monotonic reasoning with justification assertions. Their proposal is based on defining operators for approximate probabilities of a justified formula given some condition formula. Using conditional probabilities, the logic models certain aspects of defeasibile inferences with justification terms. Yet the system can neither encode the defeasibility of justification terms in their internal structure nor model defeat among reasons, to mention only some differences from our initial desiderata.

Baltag, Renne, and Smets (2012) define a justification logic in which an agent may hold a justified belief that can be compromised in the face of newly received information. The logic builds on the ideas from belief revision and dynamic epistemic logic to model examples where epistemic actions cause changes to an agent's evidence. Concerning the possibility of modelling defeaters, the logic offers two dynamic operations that change the availability of evidence in a model, namely "updates" and "upgrades" (Baltag, Renne, and Smets 2012, p. 183). Evidence obtained by updates counts as "hard" or infallible, while upgrades bring about "soft" or fallible evidence. With the use of these actions, epistemic models can represent justified beliefs being defeated, for example, by means of an epistemic action of update with hard evidence. In this way, however, the mechanism by which reasons may conflict with one another is simply being "outsourced" to an extra-logical notion of fallibility and, therefore, the logic does not directly address the ways of defeat that we formalize in this paper.

Several interesting paths could be followed in connecting the logic of default justifications with formal argumentation frameworks. Among frameworks with abstract arguments, the AFRA framework (Baroni et al. 2011) with recursive attacks offers a possibility of representing attacks to attacks. This conceptual advance can be useful in connecting default reasons to abstract arguments. Our logic could be seen as closely related to the frameworks with structured arguments, which is why connections with systems such as ASPIC+ (Prakken 2010), DeLP (García and Simari 2004), SG (Hecham, Bisquert, and Croitoru 2018) and the logic-based argumentation framework by Besnard and Hunter (2001) are still to be explored. Since each of these frameworks elaborates on the notion of defeat, a thorough comparison to our logic would shed light on their formal connections. A different logic-based perspective on argumentation frameworks is given by Caminada and Gabbay (2009) and Grossi (2010). Both papers start from the idea of studying attack graphs and formalizing notions of extensions from abstract argumentation theory using modal logic, with the former approach being proof-theoretical and the latter model-theoretical. A further interesting research venue in the field of argumentation theory is the one about the logical interpretation of *prima facie* justified assumptions in (Verheij 2003). The DefLog system which is developed there is closely related to ours in motivation, but it develops from a perspective of a sentence-based theory of defeasible reasoning instead of a rule-based or argument-based approach.

Ever since the concept of justification entered into epistemic logics, there has been a tendency to model mainstream epistemology examples, proposed by e.g. Russell, Dretske and Gettier, with the use of justification logic (Artemov 2008; 2018). With the introduction of default justifications, however, we can expect a more full-blooded integration of the formal theory of justification with the study of knowledge in philosophy, since paradigmatic examples include both incomplete specification of reasons and defeated reasons. Potential benefits of a non-monotonic system of justifications in this context were anticipated by Artemov in (2008, p. 482) where he states that "to develop a theory of non-monotonic justifications which prompt belief revision" stands as an "intriguing challenge". One of many interesting topics from epistemology that could be investigated with default-justifications theory is how does accrual of justifica-

[15] See (Fitting 2016) for a good overview of realization theorems.

tion affect the degree of justification.[16]

Appendix

Proof of Theorem 6. The claim from left to right is obvious. For the other direction, take $\mathcal{CS}$ to be some specific axiomatically appropriate and injective constant specification. We first show that if a set Γ is $\mathbf{JT_{CS}}$-finitely satisfiable, then for all formulas $F \in Fm$, it holds that $\Gamma \cup \{F\}$ or $\Gamma \cup \{\neg F\}$ is $\mathbf{JT_{CS}}$-finitely satisfiable. Suppose that Γ is $\mathbf{JT_{CS}}$-finitely satisfiable and that $\Gamma \cup \{F\}$ and $\Gamma \cup \{\neg F\}$ are both not $\mathbf{JT_{CS}}$-finitely satisfiable. Then there would be finite subsets Γ' and Γ'' of Γ such that $\Gamma' \cup \{F\}$ and $\Gamma'' \cup \{\neg F\}$ are not $\mathbf{JT_{CS}}$ satisfiable. Since for no interpretation $\mathcal{I}$ it holds that $\mathcal{I} \models \{F, \neg F\}$, $\Gamma' \cup \{F, \neg F\}$ is never $\mathbf{JT_{CS}}$ satisfiable. But since for any possible interpretation $\mathcal{I}$ one of the formulas F or $\neg F$ holds, this means that $\mathcal{I} \models \Gamma' \subseteq \mathcal{I} \models \neg F$. In a similar way we get that $\mathcal{I} \models \Gamma'' \subseteq \mathcal{I} \models F$. Therefore, we have that $\mathcal{I} \models \Gamma' \cap \mathcal{I} \models \Gamma'' = \emptyset$ and, thus, $\Gamma' \cup \Gamma''$ is not $\mathbf{JT_{CS}}$-satisfiable. But $\Gamma' \cup \Gamma''$ is a finite subset of Γ and this contradicts the assumption that Γ is $\mathbf{JT_{CS}}$-finitely satisfiable.

The next step is proving a $\mathbf{JT_{CS}}$ variant of the Lindenbaum lemma. Using the above-proven statement that for any $\mathbf{JT_{CS}}$-finitely satisfiable set of formulas Γ and any formula F, $\Gamma \cup \{F\}$ or $\Gamma \cup \{\neg F\}$ is $\mathbf{JT_{CS}}$-finitely satisfiable together with the fact that $\Gamma \cup \{F, \neg F\}$ is never $\mathbf{JT_{CS}}$-finitely satisfiable, we can construct maximally $\mathbf{JT_{CS}}$-finitely satisfiable sets. Let $F_1, F_2, F_3, \ldots$ be an enumeration of $F \in Fm$. For a $\mathbf{JT}_{CS}$-finitely satisfiable set Γ and for all $i \in \mathbb{N}$ define an increasing sequence of sets of formulas as follows:

$\Gamma_0 = \Gamma$
$\Gamma_{i+1} = \Gamma_i \cup \{F_i\}$ if $\Gamma_i \cup \{F_1\}$ is $\mathbf{JT}_{CS}$-finitely satisfiable, otherwise $\Gamma_{i+1} = \Gamma_i \cup \{\neg F_i\}$
$\Gamma' = \bigcup_{i=0}^{\infty} \Gamma_i$

We can prove that Γ' is $\mathbf{JT}_{CS}$-finitely satisfiable by induction. The base case $\Gamma_0 = \Gamma$ holds by assumption. Then we claim that for all $i \in \mathbb{N}$, Γ_i is $\mathbf{JT_{CS}}$-finitely satisfiable. For some $n \in \mathbb{N}$, take Γ_n to be $\mathbf{JT_{CS}}$-finitely satisfiable. Then either $\Gamma \cup \{F_n\}$ or $\Gamma \cup \{\neg F_n\}$ is $\mathbf{JT_{CS}}$-finitely satisfiable and, therefore, Γ_{n+1} is also $\mathbf{JT_{CS}}$-finitely satisfiable.

From the construction of the increasing sequence, we have that for any finite set $\Gamma_k \subseteq \Gamma'$ there is a $\mathbf{JT_{CS}}$-finitely satisfiable finite set $\Gamma_{k+1} \subseteq \Gamma'$ such that $\Gamma_k \subseteq \Gamma_{k+1}$ and, therefore, Γ_k is $\mathbf{JT_{CS}}$-satisfiable. Since any finite subset of Γ' is $\mathbf{JT_{CS}}$ satisfiable, Γ' is $\mathbf{JT_{CS}}$-finitely satisfiable. The set Γ' is maximal according to the enumeration of the set of formulas Fm and contains exactly one of F_i or $\neg F_i$ for all $i \in \mathbb{N}$.

Now we define a valuation v such that $v(P) = True$ iff $P \in \Gamma'$ and the reason assignment $*(t) = \{F \mid t : F \in \Gamma'\}$. We only need to check the conditions on the reason assignment function. First, we show that $*(\cdot)$ satisfies the application condition. Since the formula $t : (F \to G) \to (u : F \to (t \cdot u) : G)$ is $\mathbf{JT_{CS}}$ valid, it is contained in Γ'. If $F \to G \in *(t)$ and $F \in *(u)$, then $\{t : (F \to G), u : F\} \in \Gamma'$. Since Γ is closed under *Modus ponens*, we have that $(t \cdot u) : G \in \Gamma'$ and, therefore, $G \in *(t \cdot u)$. Similarly, since the formulas $t : F \to (t + u) : F$ and $u : F \to (t + u) : F$ are both in Γ' we can easily check that the sum condition holds for $*(\cdot)$.

Finally, we have defined an interpretation $\mathcal{I} = (*, v)$ that meets $\mathcal{CS}$ and we need to prove that truth in this interpretation is equivalent to inclusion in Γ':

$\mathcal{I} \models F$ iff $F \in \Gamma'$

The proof is by induction on the structure of F. For the base case, suppose F is an atomic formula P: $\mathcal{I} \models P$ iff $v(P) = True$ iff $P \in \Gamma'$.

For the inductive step, suppose that if the result holds for F and G, then it also holds for $\neg F$, $F \wedge G$, $F \vee G$, $F \to G$ and $t : F$. For the negation case: $\mathcal{I} \models \neg F$ iff $\mathcal{I} \not\models F$. By the inductive hypothesis, $\mathcal{I} \not\models F$ iff $F \notin \Gamma'$. By the maximality of Γ', we have that $F \notin \Gamma'$ iff $\neg F \in \Gamma'$.

For the conjunction case: $\mathcal{I} \models F \wedge G$ iff $\mathcal{I} \models F$ and $\mathcal{I} \models G$. By the inductive hypothesis, $\mathcal{I} \models F$ and $\mathcal{I} \models G$ iff $F \in \Gamma'$ and $G \in \Gamma'$ iff $F \wedge G \in \Gamma'$. Since other connectives are definable in terms of $\neg$ and $\wedge$, we skip the remaining cases.

Finally for the justified formula case: $\mathcal{I} \models t : F$ iff $F \in *(t)$. By the definition of $*(\cdot)$, it holds that $F \in *(t)$ iff $t : F \in \Gamma'$.

Therefore, for any $\mathbf{JT_{CS}}$-finitely satisfiable set Γ there is an interpretation $\mathcal{I}$ based on a maximal $\mathbf{JT_{CS}}$-finitely satisfiable extension Γ' of Γ such that $\mathcal{I} \models \Gamma$. □

[16] The question is prominent in Pollock's work (Pollock 2001).

References

Antoniou, G. 1997. *Nonmonotonic Reasoning*. Cambridge, MA: MIT Press.

Artemov, S. N., and Nogina, E. 2005. Introducing justification into epistemic logic. *Journal of Logic and Computation* 15(6):1059–1073.

Artemov, S. N. 2001. Explicit provability and constructive semantics. *Bulletin of Symbolic logic* 1–36.

Artemov, S. N. 2008. The logic of justification. *The Review of Symbolic Logic* 1(4):477–513.

Artemov, S. N. 2018. Justification awareness models. In Artemov, S. N., and Nerode, A., eds., *International Symposium on Logical Foundations of Computer Science*, volume 10703 of LNCS, 22–36. Springer.

Baltag, A.; Renne, B.; and Smets, S. 2012. The logic of justified belief change, soft evidence and defeasible knowledge. In Ong, L., and de Queiroz, R., eds., *International Workshop on Logic, Language, Information, and Computation*, 168–190. Springer.

Baltag, A.; Renne, B.; and Smets, S. 2014. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic* 165(1):49–81.

Baroni, P.; Cerutti, F.; Giacomin, M.; and Guida, G. 2011. Afra: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning* 52(1):19–37.

Besnard, P., and Hunter, A. 2001. A logic-based theory of deductive arguments. *Artificial Intelligence* 128(1-2):203–235.

Booth, R.; Meyer, T.; and Varzinczak, I. 2012. PTL: A propositional typicality logic. In del Cerro, L. F.; Herzig, A.; and Mengin, J., eds., *Logics in Artificial Intelligence: Proceedings of the 13th European conference on Logics in Artificial Intelligence*, volume 7519 of LNCS, 107–119. Springer-Verlag.

Brezhnev, V. 2001. On the logic of proofs. In Striegnitz, K., ed., *Proceedings of the Sixth ESSLLI Student Session, Helsinki*, 35–46.

Caminada, M. W., and Gabbay, D. M. 2009. A logical account of formal argumentation. *Studia Logica* 93(2-3):109.

Chisholm, R. M. 1966. *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice-Hall.

Delgrande, J. P., and Schaub, T. 2000. Expressing preferences in default logic. *Artificial Intelligence* 123(1-2):41–87.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence* 77(2):321–357.

Fan, T.-F., and Liau, C.-J. 2015. A logic for reasoning about justified uncertain beliefs. In Yang, Q., and Wooldridge, M., eds., *Proceedings of the IJCAI 2015*, 2948–2954. AAAI Press.

Fitting, M. 2005a. A logic of explicit knowledge. In Běhounek, L., and Bílková, M., eds., *Logica Yearbook 2004*. Prague: Filosofia. 11–22.

Fitting, M. 2005b. The logic of proofs, semantically. *Annals of Pure and Applied Logic* 132(1):1–25.

Fitting, M. 2008. Justification logics, logics of knowledge, and conservativity. *Annals of Mathematics and Artificial Intelligence* 53(1-4):153–167.

Fitting, M. 2009. Reasoning with justifications. In *Towards Mathematical Philosophy*. Springer. 107–123.

Fitting, M. 2016. Modal logics, justification logics, and realization. *Annals of Pure and Applied Logic* 167(8):615–648.

García, A. J., and Simari, G. R. 2004. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming* 4(1+ 2):95.

Gödel, K. 1995. Vortrag bei Zilsel/Lecture at Zilsels (1938a). In *Kurt Gödel: Collected Works: Volume III: Unpublished Essays and Lectures*, volume 3. Oxford University Press. 87–114.

Grossi, D. 2010. Argumentation in the view of modal logic. In McBurney, P.; Rahwan, I.; and Parsons, S., eds., *7th International Workshop on Argumentation in Multi-Agent Systems, ArgMAS 2010*, volume 6614 of LNCS, 190–208. Springer.

Hecham, A.; Bisquert, P.; and Croitoru, M. 2018. On a flexible representation for defeasible reasoning variants. In Dastani, M.; Sukthankar, G.; André, E.; and Koenig, S., eds., *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2018*, 1123–1131. International Foundation for Autonomous Agents and Multiagent Systems.

Horty, J. F. 2012. *Reasons as Defaults*. Oxford University Press.

Kokkinis, I.; Maksimović, P.; Ognjanović, Z.; and Studer, T. 2015. First steps towards probabilistic justification logic. *Logic Journal of the IGPL* 23(4):662–687.

Kokkinis, I.; Ognjanović, Z.; and Studer, T. 2016. Probabilistic justification logic. In Artemov, S. N., and Nerode, A., eds., *International Symposium on Logical Foundations of Computer Science*, volume 9537 of LNCS, 174–186. Springer.

Kuznets, R. 2000. On the complexity of explicit modal logics. In Clote, P. G., and Schwichtenberg, H., eds., *Computer Science Logic: 14th International Workshop, CSL 2000*, volume 1862 of LNCS, 371–383. Springer-Verlag.

Milnikel, R. S. 2007. Derivability in certain subsystems of the logic of proofs is Π_2^p-complete. *Annals of Pure and Applied Logic* 145(3):223–239.

Milnikel, R. S. 2014. The logic of uncertain justifications. *Annals of Pure and Applied Logic* 165(1):305–315.

Mkrtychev, A. 1997. Models for the logic of proofs. In Adian, S., and Nerode, A., eds., *Logical Foundations of Computer Science, 4th International Symposium, LFCS '97*, volume 1234 of LNCS, 266–275. Springer-Verlag.

Ognjanović, Z.; Savić, N.; and Studer, T. 2017. Justification logic with approximate conditional probabilities. In Baltag, A.; Seligman, J.; and Yamada, T., eds., *Logic, Rationality and Interaction, 6th International Workshop, LORI 2017*, volume 10455 of LNCS, 681–686. Springer.

Pollock, J. L. 1987. Defeasible reasoning. *Cognitive Science* 11(4):481–518.

Pollock, J. L. 2001. Defeasible reasoning with variable degrees of justification. *Artificial intelligence* 133(1-2):233–282.

Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1(2):93–124.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13(1-2):81–132.

Renne, B. 2012. Multi-agent justification logic: Communication and evidence elimination. *Synthese* 185(1):43–82.

Su, C.-P.; Fan, T.-F.; and Liau, C.-J. 2017. Possibilistic justification logic: Reasoning about justified uncertain beliefs. *ACM Transactions on Computational Logic (TOCL)* 18(2):15.

Verheij, B. 2003. DefLog: On the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation* 13(3):319–346.

Zorn, M. 1935. A remark on method in transfinite algebra. *Bulletin of the American Mathematical Society* 41(10):667–670.