From the Logic of Proofs to the Logic of Arguments

Stipe Pandžić

Theoretical Philosophy Group, Department of Philosophy and Religious Studies, Utrecht University, The Netherlands

Abstract

This paper reports on the development of default justification logic. The logic was developed as a part of the PhD project [Pandžić, 2020b], with an aim to model defeasible arguments as formulas of the justification logic object-language. The resulting logic is unique for its potential to determine the acceptability of arguments by using only the definition(s) of logical consequence.

1 Introduction

The seminal work of Toulmin [Toulmin, 2003] gave rise to the idea that the use of formal logic in argumentation is limited to deductive arguments. Toulmin's work anticipated the trend that led to the development of "informal logic" with an aim to analyze non-deductive arguments in a natural language setting. The interest in formal methods for argumentation followed the emergence of artificial intelligence research, where Pollock's work [Pollock, 1987] on defeasible reasoning stands as a landmark.

Within this research venue, multiple formal systems have been proposed to advance the computational study of structured arguments. Among them are ABA [Bondarenko et al., 1997], deductive argumentation [Besnard and Hunter, 2001], DefLog [Verheij, 2003] and ASPIC+ [Prakken, 2010]. All the mentioned systems use logical tools to represent arguments with their internal structures (not only as atomic entities as done in [Dung, 1995]) to formalize the idea of opposition between arguments by providing a theory of argumentative attacks and to define procedures that give those arguments that are undefeated. However, none of these argumentation formalisms counts as a logical system with a definition of logical consequence.

This paper reports on the logic of default justifications (or reasons) [Pandžić, 2018, Pandžić, 2019, Pandžić, 2020a, Pandžić, 2020b, Pandžić, 2021] that not only uses formal logic notions to model arguments, but rather is itself a full-fledged logic of arguments that manipulates structured arguments at a purely symbolic level. In default justification logic, structured arguments are formalized as object-level formulas of the logical system. The PhD project [Pandžić, 2020b] is centered around this logic that responds to Toulmin's anti-formalistic misgivings by providing a method to determine whether a non-deductive argument is acceptable or not through a normative system with logical consequence.

The next section outlines the underlying method by which we represent structured arguments, starting from a novel logical theory of default reasons. To be able to talk about reasons, we rely on a logical language that goes beyond propositions and represents reasons as a part of its syntax. Such a theory of reasons is one of the fundamental contributions of justification logics [Artemov, 1995, Artemov, 2001]. Justification logics tell us that a pattern of reasoning from starting premises can be captured by means of labels preceding statements and, thereby, forming *justification assertions*:

$$reason{:}conclusion.$$

This method enables us to develop a system that features reasons as logical entities *sui generis*, and we will shortly describe it in more detail.

2 Method

We start by explaining the basics of justification logics [Artemov and Fitting, 2019]. Justification logics are modal-like logics, often included in the class of epistemic logics.[1] Informally, the idea of justification logics is to replace epistemic modalities "**K**" ("in all accessible alternatives...") with explicit justifications "t" ("for the reason t...") preceding a formula F so as to produce justification assertions $t : F$ informally read as "F is known for the reason t" or "t justifies F". In this paper, we will only use two basic operations on reason terms *Application* ('$\cdot$') and *Sum* ('+'). Application produces a reason term $(u \cdot t)$ for a formula G which is a syntactic "imprint" of the *modus ponens* step from $F \to G$ and F to G for some labelled formulas $u : (F \to G)$ and $t : F$. Sum can be understood as a method to concatenate reasons in such a way that if t is a reason for F, then we can produce the reason term $(t+u)$ which is also a reason for F.

Formally, the set of exactly all reason terms (polynomials) Tm is built from justification variables $x_1, \ldots, x_n, \ldots$ and justification constants $c_1, \ldots, c_n, \ldots$ using two operators ('+' and '$\cdot$'), according to the following grammar:

$$\boxed{t ::= x \mid c \mid (t \cdot t) \mid (t + t)}.$$

Using the grammar of reason terms, we define the set Fm of justification logic formulas built according to the following grammar:

$$\boxed{F ::= \top \mid P \mid (F \to F) \mid (F \vee F) \mid (F \wedge F) \mid \neg F \mid t : F},$$

where $P \in \mathcal{P}$ and $\mathcal{P}$ is a countable set of atomic propositional formulas and $t \in Tm$.

We use the logic **JT** first introduced in [Brezhnev, 2001]. This is the weakest logic with non-defeasible and factive (truth-inducing) reasons, that is, with the axioms A2 and A3 below. This system is then generalized to model defeasible reasons. These are the axioms and rules of **JT**:

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

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

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

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

The logic **JT** is equipped with the inference rules *Modus ponens* and *Iterated axiom necessitation*. Axiom necessitation rules ensure justification of basic logical postulates. Each axiom instance of A0-A3 is justified by a justification constant, and the iterated version ensures that such justified formulas are themselves justified by a justification constant.[2]

The interpretation $\mathcal{I}$ for **JT** relies on two functions. Together with a standard *truth assignment* function for propositional formulas in $\mathcal{P}$, $\mathcal{I}$ relies on a *reason assignment* $*(\cdot) : Tm \to 2^{Fm}$, which is a function mapping each term to a set of formulas from Fm. For an interpretation $\mathcal{I}$, $\models$ is a truth relation on the set of formulas of **JT** with the following clause for a justification assertion $t : F \in Fm$: $\mathcal{I} \models t : F$ iff $F \in *(t)$. Thus, in the basic logic **JT**, the connection between a reason t and a formula F is given by means of a reason assignment function. But in our default semantics, whether t justifies F or not has to be decided on the basis of other available justifications that might defeat t. Instead of a fully specified function $*(\cdot)$, the main task of the logic of default reasons is to give a procedure to determine if t should be accepted as a justification for F.

We start by defining default theories and default rules with **JT** formulas. A default theory T is a pair (W, D), where the set W is a finite set of **JT** formulas and D is a countable set of default rules. Each default rule is of the following form:

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

The informal reading of the default δ is: "If t is a reason justifying F, and it is consistent to assume that

[1] Historically, the first justification logic was the logic of formal proofs in arithmetic **LP**, which was fully developed in [Artemov, 2001].

[2] In justification logics, justifying logical truths usually takes into account multiple restrictions and conditions that we omit here [Artemov and Fitting, 2019, pp. 17-18].

$(u \cdot t)$ is a reason justifying G, then $(u \cdot t)$ is a defeasible reason justifying G".[3] We can think about the strategy of introducing uncertain information via δ as analogous to the strategy of introducing uncertain information via Reiter's default rules [Reiter, 1980]. The main difference is that Reiter's defaults build on first-order formulas and our defaults build on justification assertions.

We exemplify the above defined notions of default rules and default reasons by means of (an extension of) an example taken from [Toulmin, 2003, p. 103].

Example Toulmin. Take S to be the proposition "Petersen is Swedish" and take R to be the proposition "Petersen is a Roman Chatolic". For some specific individual justifications p and q, δ' describes the default reasoning encoded by the term $(q \cdot p)$:

$$\delta' = \frac{p : S :: (q \cdot p) : \neg R}{(q \cdot p) : \neg R}.$$

Informally, δ' reads as follows: "If p is a reason justifying that Petersen is Swedish and it is consistent for you to assume that this gives you a reason $(q \cdot p)$ justifying that Petersen is not a Roman Catholic, then you have a defeasible reason $(q \cdot p)$ justifying that Petersen is not a Roman Catholic". Were you then to learn that although Petersen is a Swede, "Petersen's parents emigrated from Austria" (A), you would have an *undercutting defeater* (exclusionary reason) for $(q \cdot p)$:

$$\delta'' = \frac{r : A :: (s \cdot r) : \neg[q : (S \to \neg R)]}{(s \cdot r) : \neg[q : (S \to \neg R)]},$$

for some specific justifications s and r. The consequent of the rule δ'' is read as "you have a defeasible reason $(s \cdot r)$ denying that the reason q justifies that if Petersen is Swedish, then he is not a Roman Catholic". Moreover, suppose that you recall that "Petersen went to Saint Eric's Cathedral in Stockholm this Christmas" (C). This would give you an independent *rebutting defeater* and a reason to justifying that Petersen is a Roman Catholic. For some specific justifications t and u, we define:

$$\delta''' = \frac{t : C :: (u \cdot t) : R}{(u \cdot t) : R}.$$

[3]We impose three uniqueness criteria for the reason term u introduced by δ. See in [Pandžić, 2021, §3].

The example can be described by the default theory $T_1 = (W_1, D_1)$ with $W_1 = \{p : S, r : A, t : C\}$ and $D_1 = \{\delta', \delta'', \delta'''\}$.

The semantics for our default theories is inspired by Antoniou's [Antoniou, 1997] operational semantics for Reiter's default theories. Applicability of defaults is determined in the following way: for a set of deductively 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 $\neg(u \cdot t) : G \notin \Gamma$. Default consequents are brought together by applying sequences of defaults $\Pi = (\delta_0, \delta_1, \ldots)$. In this way, we obtain the evidence base set $In(\Pi)$ that is defined as a deductive closure of the union $W \cup \{cons(\delta) \mid \delta \text{ occurs in } \Pi\}$. Intuitively, $In(\Pi)$ collects reason-based information that is yet to be determined as acceptable or unacceptable.

The focus of our semantics is on meaningful Π-sequences called "processes". 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 $\Pi[k]$ is the segment of Π listing all the rules whose application is considered before δ_k.

The richness of the language of justifications enables us to give a logical theory of the notions from structured argumentation. The function *warrant assignment* $\#(\cdot) : D \to Fm$ maps each default rule to a specific justified conditional as follows:

$$\#(\delta) = u : (F \to G),$$

where $\delta \in D$ and $\delta = \frac{t:F::(u \cdot t):G}{(u \cdot t):G}$. Inspired by Toulmininan warrants, our warrants can be seen not only as specific formulas, but also as rules or the underlying principles of default generalizations.

This twofold nature of warrants is used to define the basic type of conflict between default reasons, namely, *undercutting*. A reason u undercuts reason t being a reason for a formula F in a set of **JT**-closed formulas $\Gamma \subseteq In(\Pi)$ iff $u : \neg[v : (G \to H)] \in \Gamma$, for some subterm u of t, and $v : (G \to H) = \#(\delta)$, for some δ that is applied to a process of T. In T_1, the consequent of δ'' or, more precisely, the default reason $(s \cdot r)$ undercuts $(q \cdot p)$ as a reason for $\neg R$.

Using the definition of undercut, we can specify conflict-free sets of formulas. Formal details can be found in [Pandžić, 2021, p. 17], together with a proof that each conflict-free set defined in terms of undercut is also rebuttal-free. In our logic, rebuttal ultimately results from rule inapplicability due to **JT** inconsistency. This is the case with the rebuttal between $(q \cdot p)$ and $(u \cdot t)$ in T_1, which is induced by the contradiction between $\neg R$ and R.

The standard intuitions behind defending arguments are captured in the definition of acceptability. Informally, $t : F$ is acceptable w.r.t. a set of **JT** formulas whenever that set undercuts all the undercutters of t within a relevant evidence base set $In(\Pi)$.

We define multiple argumentation theory extensions for any default theory $T = (W, D)$ starting from *potential extensions* sets Γ that always contain formulas from W and, possibly, consequents of defaults for some process Π of T.

JT-Admissible Extension A potential extension set of **JT** formulas $\Gamma \subset In(\Pi)$ is a **JT**-admissible extension of a default theory $T = (W, D)$ iff Γ is conflict-free, each formula $t : F \in \Gamma$ is acceptable with respect to Γ and Π is closed.

JT-Preferred Extension A deductive closure of a **JT**-admissible extension Γ is a **JT**-preferred extension of T iff for any other **JT**-admissible extension Γ', $\Gamma \not\subset \Gamma'$.

JT-Stable Extension A conflict-free deductive closure of a potential extension Γ is a **JT**-stable extension iff Γ undercuts all the default consequents outside its deductive closure, for all the defaults applied to any process of T.

Other standard extensions from argumentation theory and more detailed presentation of argumentation semantics in justification logic can be found in [Pandžić, 2020b, Ch. 2] and [Pandžić, 2021, § 3].

3 Discussion

A recapitulation of T_1 is given in Figure 1. According to the definitions from Section 2, we can confirm that **JT**-stable and **JT**-preferred extension of T_1 coincide in the deductive closure of $W_1 \cup \{(s \cdot r) : \neg[q : (S \to \neg R)], (u \cdot t) : R\}$. Note that the process $\Pi = (\delta', \delta'')$ includes a revision of acceptable reasons, because $(s \cdot r)$ undercuts $(q \cdot p)$.

The logic of arguments sketched here opens up a possibility to fully formalize Toulmin's argumentative

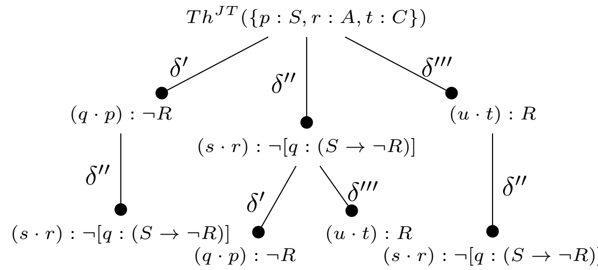

Figure 1: The process tree of T_1

schema (cf. [Pandžić, 2021, p. 16]). An important step to fully realize this is to use the first-order variant of justification logic [Fitting, 2014], instead of the propositional justification logic used here. This will enable a faithful formalization of Toulmin's warrants in the following format of default rules:

$$\frac{t_{\{x\}} : F :: (u \cdot t)_{\{x\}} : G}{(u \cdot t)_{\{x\}} : G},$$

where x in $t_{\{x\}} : F$ is a free variable throughout the derivation t. Such rules will act as default schemas that convey the generality of Toulmin's warrants.

4 Conclusion

Default justification logic is remarkable for its potential to model structured defeasible arguments as formulas of the object-language. In [Pandžić, 2020b, Ch. 3] and [Pandžić, 2021, § 4], a reader can find a method to translate between our default theories and abstract argumentation frameworks [Dung, 1995]. These methods embed our logic in the existing study of formal argumentation. Some aspects of this logic have not been mentioned here, e.g. the dynamics of default theories and *undermining* attacks as defined in [Pandžić, 2020a]. However, we at least sketch here how to "climb up" to the concepts of argumentation theory starting from a language initially proposed to represent formal proofs of mathematical statements. At the end of this paper, we indicate one of the most interesting open problems that is currently being investigated, namely, that of using first-order justification logic as a basic logic of default theories.

Acknowledgements

I am grateful to Allard Tamminga, Barteld Kooi and Rineke Verbrugge for their generous support throughout my PhD project and to the two OHAAI reviewers for their helpful suggestions on this paper. My PhD thesis was written with the support of the *Ammodo KNAW* project *Rational Dynamics and Reasoning* awarded to Barteld Kooi.

References

[Antoniou, 1997] Antoniou, G. (1997). *Nonmonotonic Reasoning.* Cambridge, MA: MIT Press.

[Artemov, 1995] Artemov, S. N. (1995). Operational modal logic. Technical Report MSI 95–29, Mathematical Sciences Institute, Cornell University.

[Artemov, 2001] Artemov, S. N. (2001). Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, pages 1–36.

[Artemov and Fitting, 2019] Artemov, S. N. and Fitting, M. (2019). *Justification Logic: Reasoning with Reasons*, volume 216 of *Cambridge tracts in mathematics.* Cambridge University Press.

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

[Bondarenko et al., 1997] Bondarenko, A., Dung, P. M., Kowalski, R. A., and Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101.

[Brezhnev, 2001] Brezhnev, V. (2001). On the logic of proofs. In Striegnitz, K., editor, *Proceedings of the Sixth ESSLLI Student Session, Helsinki*, pages 35–46.

[Dung, 1995] 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.

[Fitting, 2014] Fitting, M. (2014). Possible world semantics for first-order logic of proofs. *Annals of Pure and Applied Logic*, 165(1):225–240.

[Pandžić, 2018] Pandžić, S. (2018). A logic of default justifications. In Fermé, E. and Villata, S., editors, *Nonmonotonic Reasoning, 17th International Workshop, NMR 2018*, pages 126–135.

[Pandžić, 2019] Pandžić, S. (2019). Reifying default reasons in justification logic. In Beierle, C., Ragni, M., Stolzenburg, F., and Thimm, M., editors, *Proceedings of the KI 2019 Workshop on Formal and Cognitive Reasoning, DKB-KIK 2019*, volume 2445, pages 59–70. CEUR Workshop Proceedings.

[Pandžić, 2020a] Pandžić, S. (2020a). On the dynamics of structured argumentation: Modeling changes in default justification logic. In Herzig, A. and Kontinen, J., editors, *Foundations of Information and Knowledge Systems, 11th International Symposium, FoIKS 2020*, volume 12012 of LNCS, pages 222–241. Springer.

[Pandžić, 2020b] Pandžić, S. (2020b). *Reasoning with Defeasible Reasons.* PhD thesis, University of Groningen.

[Pandžić, 2021] Pandžić, S. (2021). A logic of defeasible argumentation: Constructing arguments in justification logic. *Argument & Computation*, Pre-press:1–45.

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

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

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

[Toulmin, 2003] Toulmin, S. E. (1958/2003). *The Uses of Argument.* Cambridge University Press.

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