Artificial Intelligence 262 (2018) 15–51

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Classical logic, argument and dialectic

M. D'Agostino [a], S. Modgil [b,*]

[a] Department of Philosophy, University of Milan, Italy
[b] Department of Informatics, King's College London, United Kingdom

ARTICLE INFO

Article history:
Received 10 July 2017
Received in revised form 10 May 2018
Accepted 24 May 2018
Available online xxxx

Keywords:
Classical logic
Argumentation
Dialectic
Rationality postulates
Natural deduction
Preferred subtheories
Bounded reasoning

ABSTRACT

A well studied instantiation of Dung's abstract theory of argumentation yields argumentation-based characterisations of non-monotonic inference over possibly inconsistent sets of classical formulae. This provides for single-agent reasoning in terms of argument and counter-argument, and distributed non-monotonic reasoning in the form of dialogues between computational and/or human agents. However, features of existing formalisations of classical logic argumentation (*Cl-Arg*) that ensure satisfaction of rationality postulates, preclude applications of *Cl-Arg* that account for real-world dialectical uses of arguments by resource-bounded agents. This paper formalises *dialectical classical logic argumentation* that both satisfies these practical desiderata and is provably rational. In contrast to standard approaches to *Cl-Arg* we: 1) draw an epistemic distinction between an argument's premises accepted as true, and those assumed true for the sake of argument, so formalising the dialectical move whereby arguments' premises are shown to be inconsistent, and avoiding the *foreign commitment* problem that arises in dialogical applications; 2) provide an account of *Cl-Arg* suitable for real-world use by eschewing the need to check that an argument's premises are subset minimal and consistent, and identifying a minimal set of assumptions as to the arguments that must be constructed from a set of formulae in order to ensure that the outcome of evaluation is rational. We then illustrate our approach with a natural deduction proof theory for propositional classical logic that allows measurement of the 'depth' of an argument, such that the construction of *depth-bounded* arguments is a tractable problem, and each increase in depth naturally equates with an increase in the inferential capabilities of real-world agents. We also provide a resource-bounded argumentative characterisation of non-monotonic inference as defined by Brewka's Preferred Subtheories.

1. Introduction

Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. While informal studies of argumentation have a rich tradition, recent years have witnessed intensive study of logic-based models of argumentation and their use in formalising agent reasoning, decision making, and inter-agent dialogue [11,53]. Much of this work builds on Dung's seminal theory of abstract argumentation [26], and the theory's provision of argumentative characterisations of nonmonotonic inference. Given a possibly inconsistent set of logical formulae (*base*) one defines the arguments and a binary *attack* relation denoting that one argument is a

* Corresponding author.
E-mail address: sanjay.modgil@kcl.ac.uk (S. Modgil).

https://doi.org/10.1016/j.artint.2018.05.003
0004-3702/

counter-argument to another. Various developments of Dung's theory additionally accommodate a preference relation over arguments, which is used to determine which attacks succeed as defeats [1,9,43]. The resulting directed graph of arguments related by defeats, referred to as an *argumentation framework* (*AF*), is then said to be 'instantiated' by the base. Evaluation of the justified arguments is then based on the intuitive principle that an argument is justified if all its defeaters are themselves defeated by justified arguments. The conclusions of justified arguments identify the 'argumentation defined' non-monotonic inferences from the instantiating base.

The widespread impact of Dung's theory is in large part due to this characterisation of non-monotonic inference in terms of the dialectical use of arguments and counter-arguments familiar in everyday reasoning and debate. The theory thus provides foundations for reasoning by individual computational and human agents, and distributed non-monotonic reasoning involving agents resolving conflicts amongst beliefs or deciding amongst alternative actions, or negotiating allocations of resources (e.g., [2,3,8,38,40,44,45,48,58]). These 'monological' and 'dialogical' applications have motivated the study of rationality postulates for logical instantiations of *AF*s [15,16], as well as desiderata for practical applications [27,44].

This paper focuses on classical logic instantiations of *AF*s (*Cl-Arg*) [1,33,43]. Features of the current paradigm have been shown to provide sufficient conditions for satisfaction of the rationality postulates. However, these features preclude satisfaction of practical desiderata that account for modelling real-world uses of arguments by resource-bounded agents. This paper therefore aims at an account of *Cl-Arg* that satisfies both practical desiderata *and* the rationality postulates.

In Section 2 we review Dung's theory, *Cl-Arg*, and the rationality postulates. In Section 3 we argue that monological and dialogical applications of Dung's theory require formalisation of real-world uses of argument suitable for resource-bounded agents. However, current approaches to *Cl-Arg* tacitly assume that *all* arguments defined by a base can be constructed and included in an *AF*, and that prior to inclusion the legitimacy of each constructed argument is verified by checking that its premises are consistent and not redundant in the strong sense that their conclusion is not entailed by any proper subset of the premises. These assumptions are computationally unfeasible (even in the propositional case) for real-world uses of argument by resource-bounded agents. However, they are proposed as sufficient conditions for satisfaction of the *consistency* and *closure* postulates [15] for first order *Cl-Arg* with preferences [43],[1] and of the 'non-contamination' postulates [16] for propositional *Cl-Arg* without preferences. Moreover, checking the legitimacy of arguments prior to inclusion in an *AF* fails to account for real-world uses of argument. *Firstly*, in real-world uses the inconsistency of arguments' premises is typically demonstrated dialectically. *Secondly*, agents do not interrogate premises for subset minimality. Rather, it is the specific proof-theoretic means for constructing arguments that determines whether or not premises are redundantly used in deriving the conclusion; that is, redundant in the obvious sense that they are syntactically disjoint from the remaining premises and the conclusion.

Section 3 then presents a new account of first order *Cl-Arg* that satisfies practical desiderata. Our approach introduces a new notion of argument that distinguishes amongst the premises accepted as true and those supposed true 'for the sake of argument'. We can therefore model a ubiquitous feature of dialectical *practice*, whereby the inconsistency of a set of premises Γ is shown *dialectically*, by defeats from arguments that claim that a contradiction is implied if one supposes (for the sake of argument) the truth of Γ. The distinction also solves the so called *foreign commitment* problem that arises in dialogical applications when agents are forced to commit to the premises of their interlocutors in order to challenge their arguments [17]. We also drop the computationally demanding checks on the legitimacy of arguments, and define 'partially instantiated' *AF*s that include subsets of the arguments defined by a base. We thus accommodate real-world uses of argument in which agents do not (or may not have sufficient resources to) construct all arguments from a base when determining whether arguments are justified. We show that our account satisfies standard properties of Dung's theory. We also show that despite dropping the legitimacy checks on arguments and making minimal assumptions as to the arguments defined by a base for inclusion in an *AF*, the consistency and closure postulates are satisfied (where the latter are adapted to account for the fact that not all defined arguments may be included in the *AF*). Moreover, in contrast with [43], these postulates are satisfied assuming *any* preference relation over arguments.

Finally, in Section 3 we identify the notion of an argument whose use of obviously redundant (in the sense described above) premises can be excluded proof-theoretically, in contrast with the use of impractical subset-minimality checks We generalise the 'non-contamination' postulates defined for propositional instantiations of *AF*s in [16], to first order instantiations. We then show that despite dropping consistency and subset minimality checks on arguments' premises, our account of first order *Cl-Arg* satisfies these postulates under the assumption that preference relations are 'coherent'.

Standard accounts of *Cl-Arg* typically leave implicit the specific proof theoretic means by which one entails a conclusion from a set of premises. In Section 4 we illustrate use of our dialectical account of argumentation by formalising arguments as *intelim* trees: a new natural deduction formalism for propositional classical logic [20,21] that allows measurement of the 'depth' of an argument such that the construction of *depth-bounded* arguments is a tractable problem, and each increase in depth naturally equates with an increase in the inferential capabilities of real-world agents. We then show that *AF*s instantiated by arguments up to any given depth satisfy the rationality postulates and practical desiderata. Furthermore, *intelim* natural deduction allows for a notion of proof that excludes arguments that use obviously redundant premises. Section 4 also develops a resource-bounded argumentative characterisation of non-monotonic reasoning in Brewka's Preferred Subtheories [13].

[1] Consistency postulates closely related to [15] are also studied for propositional *Cl-Arg* without preferences in [33].