Argument, Discussion and Rationality

Martin Caminada

University of Aberdeen

When applying argumentation theory for purposes of what to believe or what to do, the idea is to follow a three-step process. In the first step, one starts with a particular knowledge base and determines what are the possible defeasible derivations (called *arguments*) one can make using this knowledge base. These derivations then become the nodes of a directed graph called an *argumentation framework*. The edges of such a graph represent the *attack*-relation. The idea is that an argument A attacks an argument B iff what A derives somehow invalidates the derivation B. Once the argumentation framework has been constructed, the second step is to determine which of the arguments to accept. The fundamental idea of Dungs theory [1] is to determine this based purely on the structure of the graph, without looking at the actual contents of the arguments. Different selection criteria (called *argumentation semantics*) have been stated, often allowing for more than one possible set of accepted arguments. After determining the set (or sets) of accepted arguments, the third step in the argumentation process is to determine the set (or sets) of accepted *conclusions*. This can be done in a fairly straightforward way. For each accepted argument, the conclusion supported by its derivation will be an accepted conclusion.

Formalisms like ABA [2] and ASPIC [3] assume that the information in the knowledge-base consists of derivation rules, where ASPIC also distinguishes between defeasible and non-defeasible (strict) derivation rules. Premises can be represented as (strict) rules with an empty antecedent. Arguments are then constructed by chaining the rules together, basically in a tree-like structure, with the premises (and assumptions) being the leaf nodes, and the conclusion of the argument being the consequent of its top-rule.

A question worthwhile examining is to what extent the argumentation process can be considered to be rational. The current literature in formal argumentation allows for three approaches of examining rationality: on the semantical level (step 2) on the level of argument-based discussion, and on the level of the actual outcome of the argumentation process (step 3).

As for the semantical level, we recall that the idea is, given a graph, to determine which nodes (arguments) are to be accepted and which nodes (arguments) are to be rejected. In essence, one labels each node of the graph with either `accepted`, `rejected` or `undecided`. Although this allows one to express any arbitrary position on which arguments are accepted, rejected or undecided, some of these positions can be considered as more reasonable than others. The role of the argumentation semantics is to determine which positions can be considered as reasonable. This is done by defining constraints on the possible argument labellings. One such constraint (called *complete semantics* [1, 4]) is as follows:

- Each argument that is `accepted` has all its attackers `rejected`
- Each argument that is `rejected` has at least one attacker that is `accepted`
- Each argument that is `undecided` does not have all its attackers `rejected` and does not have an attacker that is `accepted`

The third condition says that in order for an argument to be `undecided` one does not have enough grounds to make it `accepted` (i.e. not all its attackers are `rejected`) and one does not have enough grounds to make it `rejected` (i.e. it does not have an attacker that is accepted).

It is not difficult to see the intuitive appeal of the above mentioned criteria of complete semantics. After all, given that an attack stands for the attacking argument (if it is sufficiently in force) somehow invalidating the attacked argument, it would be difficult to imagine how for instance an argument can be accepted if it also has an attacker that is accepted. However, it is important to keep in mind that the "rationality" of complete semantics (as well as its adaptions of grounded, preferred, stable, semi-stable, ideal and eager semantics) is based on its intuitive appeal, and not on an external "objective" criterion regarding the outcome.

The second approach for examining rationality is on the level of argument-based discussion. Here, the idea is to reframe the outcome of the semantics in terms of rational discussion. Grounded semantics,[1] for instance, can be captured by the *Grounded Persuasion Game* [5]. That is, an argument is `accepted` in the grounded labelling iff in a proponent-opponent discussion under the rules of the Grounded Persuasion Game, the proponent can win the discussion. Hence, the Grounded Persuasion Game can be seen as a proof procedure for grounded semantics. Different argumentation semantics correspond to different *types* of argument discussion games. Grounded semantics, for instance, corresponds with Mackenzie-style persuasion, whereas (credulous) preferred semantics corresponds with Socratic discussion [6]. Hence, a semantics can be considered as "rational" iff it corresponds with a particular type of natural discussion that is somehow also considered to be rational.

The third approach for determining rationality is by looking at the actual *outcome* of the argumentation process in terms of accepted conclusions. At the third step of the argumentation process, an argument labelling is converted to a conclusion labelling by labelling a conclusion as `accepted` iff there is an argument for the conclusion that is `accepted` and labelling a conclusion as `rejected` iff all arguments for the conclusion are `rejected` [7]. One of the desirable properties one would like to have is that the set of all accepted conclusions is classically consistent. Furthermore, one would also like the accepted conclusions to be closed under the strict rules (e.g. if "Socrates is a man" is an accepted conclusion, and "Each man is mortal" is a strict rule in the knowledge base, then "Socrates is mortal" should also be an accepted conclusion). Principles like these are known as *rationality postulates* [3] and much research has been carried out to satisfy them. The key difficulty to be overcome is that argument selection (step 2) is done in a "blind" way, without looking at what is actually inside of the arguments, and purely based on the topology of the graph. It turns out that satisfying the rationality postulates requires above all a careful combination of how to construct the graph (step 1) and how to evaluate the graph (step 2).

From the above discussion, it is clear that the current way of examining rationality in the context of formal argumentation has its shortcomings. Rationality on semantical level (including on discussion level) critically depends on its intuitive appeal, and is not based on any external criterion (like being "optimal" in any kind of way). Rationality on the conclusion level, on the other hand, is relatively weak in the sense that it is merely aimed at preventing outcome that is absurd, rather than warrenting outcome that is somehow "optimal" (hence, it should be seen as no more than a *necessary* requirement). How precisely to embed stronger notions of rationality into formal argumentation theory is still an open research question.

References

1. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77** (1995) 321–357
2. Dung, P., Kowalski, R., Toni, F.: Assumption-based argumentation. In Simari, G., Rahwan, I., eds.: Argumentation in Artificial Intelligence. Springer US (2009) 199–218
3. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artificial Intelligence **171**(5-6) (2007) 286–310
4. Caminada, M., Gabbay, D.: A logical account of formal argumentation. Studia Logica **93**(2-3) (2009) 109–145 Special issue: new ideas in argumentation theory.
5. Caminada, M., Podlaszewski, M.: Grounded semantics as persuasion dialogue. In Verheij, B., Szeider, S., Woltran, S., eds.: Computational Models of Argument - Proceedings of COMMA 2012. (2012) 478–485
6. Caminada, M., Dvořák, W., Vesic, S.: Preferred semantics as socratic discussion. Journal of Logic and Computation (2014) (in print).
7. Wu, Y., Caminada, M.: A labelling-based justification status of arguments. Studies in Logic **3**(4) (2010) 12–29

[1] The grounded labelling is the unique complete labelling where the set of `accepted` arguments is minimal. Hence, an argument is `accepted` in the grounded labelling iff it is `accepted` in every complete labelling.