

On the Relative Nature of Argument Strength

Antonis Kakas

Department of Computer Science, University of Cyprus

Email: antonis@ucy.ac.cy

The essential role of the strength of an argument is that it allows an argument to qualify as a defense against some other argument that is contesting the argument at hand or contesting an argument with which the argument at hand is in coalition. In effect, this means that what is important is the *relative strength* of an argument with respect to other arguments rather than some absolute level of an individual argument in isolation. In some cases we may want to ascribe individual strength to arguments in isolation, e.g. when we are in the process of learning arguments and we can not be a-priori sure of their value. The learning data may be incomplete and/or uncertain and hence we want to judge the individual strength or quality of the arguments as these are produced, as for example it is studied in the context of Bayesian Argumentation [4]). Nevertheless, when we come to use these arguments their individual strength is effectively used in order to lead to a relative strength between arguments on opposing sides of the dispute and the dialectic argumentation process between them.

An example of the relative nature of strength of arguments is the case of logical reasoning via argumentation, where arguments that are based on the given premises under which we are reasoning are stronger than arguments that are based on hypothetical premises outside the given theory. Hence in Logic Programming [5] an argument composed of a set of rules that derive a positive conclusion, p , is considered stronger than a “hypothetical” argument supporting directly the opposite negative conclusion, i.e. supporting the default negation as failure of *not* p . Similarly, this is the case with most, if not all, reformulations of non-monotonic logics in terms of argumentation [1].

Furthermore, this also holds for classical deductive reasoning where again one of the essential elements of the reformulation [8] of classical logical entailment in Argumentation Logic is the relative strength of arguments based on the premises over those based on formulae not in the given theory. But in this case a second element of argument relative strength is needed. This is the *temporary strength* of an argument where during the dialectic reasoning this argument is assumed to be as strong as any other argument based on the premises. With this extra element we can then recover the indirect proofs of classical logic via Reductio ad Absurdum (*RAA*) by ascribing to the argument supporting the formula that we are hypothesizing in the application of reductio ad absurdum, this temporal relative strength over other (opposing) hypothetical arguments.

To properly account for this relativistic nature of strength it is necessary to relativize the semantics of argumentation and consider a relative notion of acceptability between (sets of) arguments generalizing the absolute notion of acceptability as it is normally found for example in abstract argumentation [3]. In this, the dialectic argumentation semantics is defined via a relation $ACC(\Delta, \Delta_0)$

between any two sets of arguments Δ, Δ_0 which specifies the acceptability of the set of arguments Δ under the context where the set Δ_0 of arguments is considered as given, i.e. temporally as strong as any other argument and hence acceptable. This is a kind of *relative labeling* semantics [2] where the dialectic coherence of the labeling is applied not in just one absolute case but in all possible cases where some subset of arguments is assumed given and thus relatively as strong as other arguments.

This relative acceptability semantics has been studied from the early 1990s particularly in the context of Logic Programming [5, 7, 9] and more recently in the abstract setting of argumentation [6]. It constitutes a conservative extension of the widely used admissibility semantics for argumentation. Informally, the relative acceptability of $ACC(\Delta, \Delta_0)$ is defined recursively to hold when the argument set Δ can render all its attacking (or counter-arguments) non-acceptable in the context of accepting Δ_0 together with Δ (see [6] for the technical details). In practice, the essential new element that it introduces is a semantical notion of *self-defeating* arguments and how these then identify additional acceptable arguments that can not be captured by other semantics.

Informally, a self-defeating argument “turns on itself” by rendering one of its attacking arguments acceptable when the argument is temporarily considered strong enough to be accepted. This means that the self-defeating argument renders the arguments that it needs for its defence, against some attacking counter-argument, non-acceptable. More formally, we can define a self-defeating argument S as one for which there exists a counter-argument A such that $\neg ACC(A,)$ and $ACC(A, S)$ hold. So, although the attack A is in general (i.e. when we do not take any argument to be temporarily strong) non-acceptable under S temporarily strong this attack is rendered acceptable. Hence S brings about its own defeat and non-acceptability. The simplest example of a self-defeating argument is one that attacks itself, where since the argument is attacked by itself it is rendering this attack as acceptable in its own temporary context of strength. Hence in many semantics this is excluded explicitly by requiring that argument sets are not self-attacking.

For a more elaborate example of a self-defeating argument let us consider, within the framework of abstract argumentation, the case of a framework where we have three arguments a_1, a_2, a_3 and where a_1 is attacked strongly by a_2 , a_2 is attacked strongly by a_3 and a_3 is attacked by a_1 . We have therefore a three-loop in the attack relation where a_1 cannot attack back and defend itself against the attack from a_2 . For this it needs the argument a_3 as a defense against a_2 but a_3 is attacked by a_1 . Hence a_1 renders its required defense non-acceptable and thus indirectly itself non-acceptable. As a consequence of this a_1 cannot compromise the acceptability of other arguments that it is attacking. Hence, if we had in the above framework an additional argument, a_0 , that is attacked by a_1 the argument a_0 will be acceptable, despite the fact that a_0 does not counter-attack back its attack by a_1 and therefore a_0 cannot be admissible, because its (only) attack of a_1 is self-defeating.

There are two important links of self-defeating arguments defined in this way through the relative acceptability semantics. The first one is that this can be seen as a generalization or abstraction of the Reduction ad Absurdum reasoning principle, which as mentioned above, allows us to reformulate classical logical reasoning in terms of argumentation [8]. The temporary strength of an argument under which the self-defeating arguments are defined is analogous to the the process of proof by contradiction where the assumed hypothesis is given temporarily equal strength as the premises in the theory T under which we are reasoning.

The other (related) link is that self-defeating arguments are related to fallacious arguments and reasoning. They are indeed fallacious arguments that are structurally so and independent of the content or context of reasoning as the above example illustrates. They result purely from the logical structure of argumentation given by its relative acceptability semantics. This is in contrast to informal fallacies, as observed within human reasoning, which are typically context (e.g. audience) dependent fallacies and depend on the relative strength of hypothetical arguments within the context of reasoning.

Let us illustrate this case of an informal fallacy with an example. Informal fallacies can be understood in terms of arguments where the fallacious reasoning is based on some argument scheme. Consider the fallacy of “reasoning from ignorance” or “Argument from ignorance (Argumentum ad Ignorantiam)” where the argument and reasoning from it is based on lack of evidence. Consider a case of this “argument from ignorance fallacy” given by:

“ a_1 : Ghosts exist, because no one has proved that they do not exist.”

This is an argument supporting the claim of the existence of ghosts based on the premise that no one has shown otherwise. There are several ways to view this as a fallacious argument. One such view is that this link between the premises and claim is not strong, e.g. in Bayesian Argumentation this strength depends on the conditional probability of beliefs, in the population, between the conclusion and premises and that in this case this is low.

Another way to understand the fallacious nature of (the use of) this argument is through its relationship with the *hypothesis argument* for the opposite claim, i.e. the hypothesis that ghosts do not exist. This hypothesis argument is typically considered stronger than this argument and so it defeats it (it cannot be defended against), thus turning a_1 as non-acceptable. In other words, in the context or the audience in which it is applied this is a fallacious argument because of the belief in the opposite claim that is stronger not simply than the claim itself but stronger than the link to the claim by this argument: it is the argument that is defeated not simply the claim.

To emphasize the fact that this fallacy is linked to the external beliefs in the environment and it is a matter of contextual relative strength between the argument and such hypothetical arguments let us consider another specific case of the same “argument from ignorance” fallacy, exactly as the above argument a_1 on ghosts but with ghosts replaced by God. Then the strong belief that someone might have in the hypothesis or statement that “God exists” helps de-

find the argument from ignorance from any counter-argument against it because this argument of belief would be considered relatively at least as strong as the counter-argument. Thus although a_1 may be relatively weak the belief argument is relatively strong and this supports a_1 to be an acceptable argument within a coalition with the belief argument. For the particular reasoners the argument from ignorance is non-fallacious, irrespective of whether they would use it to support the claim that God exists or not.

Hence, under the relativistic acceptability semantics we can distinguish informal fallacies from other purely structural fallacies as two cases of non-acceptable arguments. In the first case the argument set contains a hypothetical argument that is relatively weak in a particular context of reasoning, whereas in the second case the argument (or argument set) is self-defeating, i.e. the adoption of this as a relatively strong argument would render it non-acceptable.

In informal common sense human reasoning allowing hypothetical arguments with varying relative strength is important in capturing human biases in the reasoning. These individual biases, or subjective beliefs, are reflected by the relative strength between one hypothesis argument and other contesting hypothesis arguments. *Cognitive Argumentation* [10] aims to study this by synthesizing together computational argumentation from AI with cognitive principles born out of empirical and theoretical findings of Cognitive Science. A related system, called *COGNICA*, models human conditional reasoning and is used to carry out experiments to evaluate argumentation as a basis for human reasoning.

References

1. A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.
2. Martin W. A. Caminada and Dov M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.
3. P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games. *Artif. Intell.*, 77:321–357, 1995.
4. U. Hahn and J. Hornikx. A normative framework for argument quality: argumentation schemes with a Bayesian foundation. *Synthese*, 193:1833–1873, 2016.
5. A. Kakas, R. Kowalski, and F. Toni. Abductive Logic Programming. *Journal of Logic and Computation*, 2(6):719–770, 1992.
6. A. Kakas and P. Mancarella. On the Semantics of Abstract Argumentation. *Journal of Logic and Computation*, 23(5):991–1015, 2013.
7. Antonis C. Kakas, Paolo Mancarella, and Phan Minh Dung. The acceptability semantics for logic programs. In *Proc. of 11th Int. Conf. on Logic Programming*, pages 504–519, 1994.
8. Antonis C. Kakas, Paolo Mancarella, and Francesca Toni. On argumentation logic and propositional logic. *Studia Logica*, 106(2):237–279, 2018.
9. Antonis C. Kakas and Francesca Toni. Computing argumentation in logic programming. *J. Log. Comput.*, 9(4):515–562, 1999.
10. Emmanuelle-Anna Dietz Saldanha and Antonis C. Kakas. Cognitive argumentation for human syllogistic reasoning. *Künstliche Intell.*, 33(3):229–242, 2019.