Towards a logic of argumentation

Leila Amgoud and Henri Prade

IRIT, Université Paul Sabatier, 31062 Toulouse Cedex 9, France
{amgoud,prade}@irit.fr

Abstract. Starting from a typology of argumentative forms proposed in linguistics by Apothéloz, and observing that the four basic forms can be organized in a square of oppositions, we present a logical language, somewhat inspired from generalized possibilistic logic, where these basic forms can be expressed. We further analyze the interplay between the formulas of this language by means of two hexagons of oppositions. We then outline the inference machinery underlying this logic, and discuss its interest for argumentation.

1 Introduction

In a work still largely ignored in Artificial Intelligence, the linguist Apothéloz [1] established a catalogue of argumentative forms more than two decades ago; see also [11]. In particular, he advocated the difference between statements such as "y is not a reason for concluding x" and "y is a reason against concluding x", which may be viewed as two different negative forms that are in opposition with the more simple statement "y is a reason for concluding x". It has been noticed in philosophical logic for a long time that the existence of two negation systems gives birth to a square of oppositions [10]. This has led Salavastru [12] to present a reading of Apothéloz' typology of argumentative forms in terms of square of oppositions a few years ago, and to propose a propositional logic translation of the basic argumentative statements. However, propositional logic is not rich enough for offering a representation setting for such a variety of statements.

In the following, after a brief reminder on the square of oppositions, we first reexamine Salavastru's proposal and identify several weaknesses. We first restate the argumentative square of opposition properly, and then introduce the basic elements of a kind of conditional logic language, somewhat inspired by generalized possibilistic logic [5], in which we can get a more suitable translation of the argumentative square. This square can in fact be extended into a more complete hexagon of opposition [3], which makes clear that its underlying structure is based on the trichotomy "y is a reason for concluding x", "y is a reason for concluding $\neg x$", "y is neither a reason for concluding x, nor for concluding $\neg x$". We then outline the inference machinery of the proposed logic, emphasize the difference between "y is a reason for concluding x" and "x follows logically from y", which leads to build another hexagon of opposition. We discuss the potential interest of the proposed logic for argumentation, and finally mention some possible extensions for handling nonmonotonic reasoning and graded argumentative statements in the spirit of possibilistic logic.

2 Argumentative square

Apothéloz [1] points out the existence of four basic argumentative forms:

- i) "y is a reason for concluding x" (denoted $\mathcal{C}(x) : \mathcal{R}(y)$),
- ii) "y is not a reason for concluding x" ($\mathcal{C}(x) : -\mathcal{R}(y)$),
- iii) "y is a reason against concluding x" ($-\mathcal{C}(x) : \mathcal{R}(y)$), and
- iv) "y is not a reason against concluding x" ($-\mathcal{C}(x) : -\mathcal{R}(y)$).

Interestingly enough, several of these forms have not been considered in Artificial Intelligence research. As can be seen several forms of opposition are present in these statements, where two negations are at work. A key point in this categorization is indeed the presence of two kinds of negation, one pertaining to the contents x or y, and the other to the functions $\mathcal{R}$ or $\mathcal{C}$. It has been observed that such a double system of negations gives birth to a formal logical structure called *square of opposition*, which dates back to Aristotle's time (see, e.g., [10] for a historical and philosophical account). We first briefly recall what this object is, since it has been somewhat neglected in modern logic.

It has been noticed for a long time that a statement (A) of the form "every a is p" is negated by the statement (O) "some a is not p", while a statement like (E) "no a is p" is clearly in even stronger opposition to the first statement (A). These three statements, together with the negation of the last statement, namely (I) "some a is p", give birth to the square of opposition in terms of quantifiers $A : \forall a\ p(a)$, $E : \forall a\ \neg p(a)$, $I : \exists a\ p(a)$, $O : \exists a\ \neg p(a)$, pictured in Figure 1. Such a square is usually denoted by the letters A, I (affirmative half) and E, O (negative half). The names of the vertices come from a traditional Latin reading: **A**ff**I**rmo, n**E**g**O**). Another standard example of the square of opposition is in terms of modalities: $A : \Box r$, $E : \Box\neg r$, $I : \Diamond r$, $O : \Diamond\neg r$ (where $\Diamond r \equiv \neg\Box\neg r$). As can be seen from these two examples, different relations hold between the vertices. It gives birth to the following definition:

Definition 1 (Square of opposition). *Four statements A, E, O, I make a* square of opposition *if and only if the following relations hold:*

- *(a) A and O are the negation of each other, as well as E and I;*
- *(b) A entails I, and E entails O;*
- *(c) A and E cannot be true together, but may be false together, while*
- *(d) I and O cannot be false together, but may be true together.*

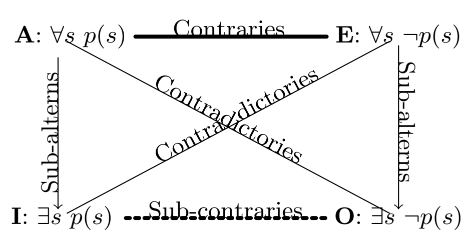

Fig. 1. Square of opposition

Note the square in Fig. 1 pressupposes the existence of some s (non empty domain). $r \not\equiv \bot, \top$ is assumed in the modal logic case.

The observation that two negations are at work in the argumentative statements classified by Apothéloz [1] has recently led Salavastru [12] to propose to organize the four basic statements into a square of opposition; see also [9]. However, his proposal is debatable on one point, as we are going to see. Indeed, taking $\mathcal{C}(x) : \mathcal{R}(y)$ for vertex A, leads to take its negation $\mathcal{C}(x) : -\mathcal{R}(y)$ for O. Can we take $-\mathcal{C}(x) : \mathcal{R}(y)$ for E? This first supposes that A and E are mutually exclusive, which is clearly the case. Then, we have to take the negation of E for I, i.e. $-\mathcal{C}(x) : -\mathcal{R}(y)$. We have still to check that A entails I and E entails O, as well as condition (d) above. If y is a reason for not concluding x, then certainly y is not a reason for concluding x, so E entails O; similarly y is a reason for concluding x entails that y is not a reason for not concluding x, i.e. A entails I. Finally, y may be a reason neither for concluding x nor for not concluding x. This gives birth to the argumentative square of opposition of Figure 2. It can be checked that the contradiction relation (a) holds, as well as the relations (b), (c), and (d) of Definition 1.

Proposition 1. *The four argumentative forms* $A = \mathcal{C}(x) : \mathcal{R}(y)$, $E = -\mathcal{C}(x) : \mathcal{R}(y)$, $O = \mathcal{C}(x) : -\mathcal{R}(y)$, $I = -\mathcal{C}(x) : -\mathcal{R}(y)$ *make a square of opposition.*

Note that we should assume that $\mathcal{C}(x) : \mathcal{R}(y)$ is not self-contradictory (or self-attacking) in order that the square of opposition really makes sense. In propositional logic, this would mean that $x \wedge y \neq \bot$.

This square departs from the one obtained by Salavastru [12] where vertices A and I as well as E and O are exchanged: In other words the entailments (b) are put in the wrong way. This may come from a misunderstanding of the remark made in [1] that the rejection $\mathcal{C}(x) : -\mathcal{R}(y)$ is itself a reason for not concluding x, which can be written $-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$. But this does not mean that $\mathcal{C}(x) : -\mathcal{R}(y)$ entails $-\mathcal{C}(x) : \mathcal{R}(y)$ since it may be the case, for instance, that $\mathcal{C}(-x) : \mathcal{R}(y)$. Salavastru made another similar mistake regarding the link between A and I. He assumed that I entails A. It can be seen on a simple example that this implication is false, and is rather in the other way round: "The fact that Paul is a French citizen (fr) is not a reason for not concluding that he is smart (st). This is clearly a statement of the form $-\mathcal{C}(sm) : -\mathcal{R}(fr)$. The question now is: does this statement entail the argument $\mathcal{C}(sm) : \mathcal{R}(fr)$ (i.e. the fact that Paul is French is a reason to conclude that he is smart)? The answer is certainly no. However, the converse is true. That is $\mathcal{C}(sm) : \mathcal{R}(fr)$ implies $-\mathcal{C}(sm) : -\mathcal{R}(fr)$.

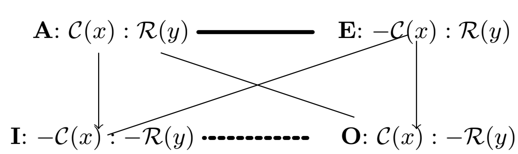

Fig. 2. An informal, argumentative square of opposition

Salavastru [12] also proposed a propositional logic reading of the informal square of opposition of Figure 2. This square is in terms of logical binary connectives and is pictured in Figure 3 (where $\uparrow$ here denotes Sheffer's incompatibility operator, which corresponds to the negation of the conjunction, i.e. $y \uparrow x = \neg y \vee \neg x$). In Salavastru's view "y is a reason for concluding x" is modeled by $y \rightarrow x$, which corresponds to a strong reading of the consequence relation. Then "y is a reason for not concluding x" is understood as the incompatibility of y and x, while "y is not a reason for not concluding x" is just their conjunction, i.e. two symmetrical connectives w.r.t. x and y, which may seem troublesome. Still from a formal point of view, this makes a perfect square of opposition. Indeed $y \wedge x$ entails $y \rightarrow x$ (and $y \not\rightarrow x$ entails $y \uparrow x$), but as already said, "y is not a reason for not concluding x" does not entail "y is a reason for concluding x" (and modeling "y is a reason for concluding x" by the symmetrical formula $y \wedge x$ would look strange). In fact, propositional logic is not enough expressive for providing a logical language for reasoning about arguments.

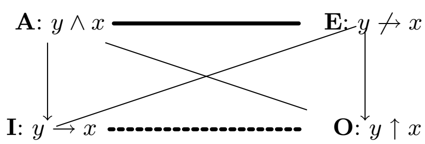

Fig. 3. Salavastru's logical square of opposition for argumentation

3 Towards a logical language for argumentative reasoning

Let x, y, z, x', y', ... denote any propositional logic formula. The basic building block of the proposed logical language is made of pairs of the form (x, y) to be read "y is a reason for x", or "x is supported by y". It will stand for $\mathcal{C}(x) : \mathcal{R}(y)$. In fact, this may be viewed as a formula of the logic of supporters [8], a counterpart of possibilistic logic [4], where the certainty level (usually belonging to an ordered chain) of a proposition x is replaced by its support (belonging to a Boolean lattice of propositions). The logic of supporters is a lattice-based generalization of possibilistic logic. See [8] for a detailed account of its semantics. In particular, if (x, y) and (x, y') hold, $(x, y \vee y')$ holds as well: if y and y' are reasons for concluding x, $y \vee y'$ is also a reason for concluding x.

As in standard possibilistic logic, the logic of supporters only allows for conjunctions of such pairs, and we have $(x \wedge x', y) = (x, y) \wedge (x', y)$. This means that if y is a reason for concluding x and for concluding x', then y is a reason for concluding $x \wedge x'$, and conversely. But, as immediately revealed by the kind of statements we have to handle, we need a *two layer* propositional-like language. Indeed, we need to express negations of such pairs, namely $\neg(x, y)$ to express $\mathcal{C}(x) : -\mathcal{R}(y)$. We also need disjunctions between such pairs as we are going to see, by completing the square of oppositions into an hexagon.

Indeed, it is always possible to complete a square of opposition into a hexagon by adding the vertices $Y =_{def} I \wedge O$, and $U =_{def} A \vee E$. This completion of a square of opposition was proposed and advocated by a philosopher and logician, Robert Blanché (see, e.g., [3]). It fully exhibits the logical relations inside a structure of oppositions generated by the three mutually exclusive situations A, E, and Y, where two vertices linked by a diagonal are contradictories, A and E entail U, while Y entails both I and O. Moreover $I = A \vee Y$ and $O = E \vee Y$. The interest of this hexagonal construct has been especially advocated by Béziau [3] in the recent years for solving delicate questions in paraconsistent logic modeling. Conversely, three mutually exclusive situations playing the roles of A, E, and Y always give birth to a hexagon [6], which is made of three squares of opposition: $AEOI$, $AYOU$, and $EYIU$.[1]

Definition 2 (Hexagon of opposition). *Six statements A, U, E, O, Y, I make a* hexagon of opposition *if and only if A, E, and Y are mutually exclusive two by two, and $AEOI$, $AYOU$, and $EYIU$ are squares of opposition.*

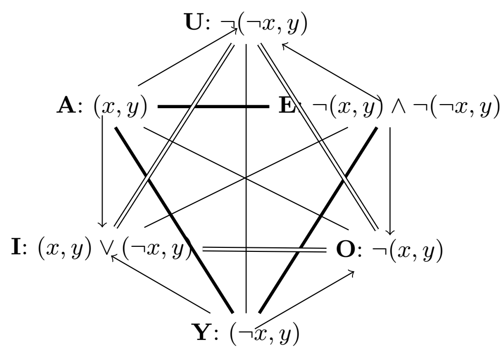

Fig. 4. Possible argumentative relations linking a reason y to a conclusion x

Figure 4 exhibits the six possible epistemic situations (apart complete ignorance) regarding argumentative statements relating y and x. Indeed it provides an organized view of the six argumentative statements, namely: A: "y is a reason for concluding x" represented by (x, y); Y: "y is a reason for concluding $\neg x$" represented by $(\neg x, y)$, O: "y is not a reason for concluding x" represented by $\neg(x, y)$; U: "y is not a reason for concluding $\neg x$" represented by $\neg(\neg x, y)$; completed by I: "y is conclusive about $x/\neg x$" represented by $(x, y) \vee (\neg x, y)$; and E: "y is not conclusive about $x/\neg x$" represented by $\neg(x, y) \wedge \neg(\neg x, y)$. Thus, matching the square of Fig. 2 with the square $AEOI$ in the hexagon of Figure 4 reveals that $-\mathcal{C}(x) : \mathcal{R}(y)$ is represented by $\neg(x, y) \wedge \neg(\neg x, y)$, i.e. "$y$ is a reason for not concluding about x" is also understood here as a reason "y is a reason for not concluding about $\neg x$".

[1] Note that, if we complete Salavastru's square of Figure 2 into a hexagon, we obtain $U = y$ and $Y = \neg y$, which corresponds to the simple affirmation and negation of y, where x is no longer involved, which is not very satisfactory.

This leads to consider a logic, we call LA, which is a propositional-like logic where *all* literals are replaced by pairs, e.g. $\neg(x, y)$, $\neg(x, y) \vee (x', y')$, $(x, y) \wedge (x', y')$ are wffs in LA (but not $\neg z \wedge (x, y)$). At the higher level, the pairs (x, y) are manipulated as literals in propositional logic, e.g., $(x, y) \wedge \neg(x, y)$ is a contradiction, and

$$(x, y), \neg(x, y) \vee (x', y') \vdash (x', y')$$

is a valid rule of inference. LA is a two-layer logic, just as the generalized possibilistic logic [5] is w.r.t. the standard possibilistic logic [4].

At the internal level, x, y are themselves propositional variables, and LA behaves as possibilistic logic: $(x, y) \wedge (x', y) \equiv (x \wedge x', y)$ as already said, and

$$(\neg x \vee x', y), (x \vee z, y') \vdash (x' \vee z, y \wedge y')$$

is a valid inference rule.

Moreover, we have: if $x \vdash x'$ then $(x, y) \vdash (x', y)$. So when $x \vdash x'$, if "y is a reason for x" then "y is a reason for x'". Writing $x \vdash x'$ as $(\neg x \vee x', \top)$, the above rule follows from the previous one: $(\neg x \vee x', \top), (x, y) \vdash (x', \top \wedge y)$ and $(x', \top \wedge y) \equiv (x', y)$. Thus $(x, y) \vdash (x \vee x', y)$. Besides, $\vdash ((x, y) \vee (x', y)) \rightarrow (x \vee x', y)$ (where $\rightarrow$ is the material implication). Note that the converse implication does not hold in general. Indeed y may be a reason for $x \vee x'$, without being a reason for x or being a reason for x'.

Note also that $(\neg x, y), (x, y') \vdash (\bot, y \wedge y')$ is a contradiction only if the reasons y and y' are *not* mutually exclusive. Generally speaking, one has to to distinguish, between

- $(\bot, y)$ (with $y \neq \bot$) which is a contradiction;
- $(x, \bot)$ (which can be obtained, e.g. from $(x \vee y, z)$ and $(\neg y, \neg z)$), which expresses that x has no support.

This should not be confused with the case where x would be equally supported by opposite reasons: (x, z) and $(x, \neg z)$ which entails that $(x, \top)$. Besides, $(\top, x)$ holds for any propositional formula $x \neq \bot$. Finally, it can be checked that:

Proposition 2. *The 6 LA formulas in Fig. 4 make a hexagon of opposition.*

Then, it can be seen in Fig. 4 that the argumentative form $\mathcal{C}(x) : -\mathcal{R}(y)$ is equal to the disjunction of the forms $\mathcal{C}(\neg x) : \mathcal{R}(y)$ and $-\mathcal{C}(x) : \mathcal{R}(y)$, which is satisfactory. Indeed $(\neg x, y)$ entails $\neg(x, y)$ (since $(x, y), (\neg x, y) \vdash (\bot, y)$).

We have only outlined how LA behaves. It is worth noticing that LA is expressive enough for allowing the use of negation in three places:

- $(x, \neg y)$ i.e., "$\neg y$ is a reason for x";
- $(\neg x, y)$, i.e., "y is a reason for $\neg x$", and
- $\neg(x, y)$ "y is not a reason for x", i.e., in particular it is possible that $\neg x$ holds while y holds.

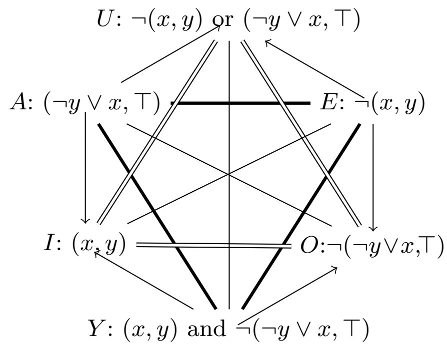

Fig. 5. Hexagon showing the interplay between a strong and a weak argumentative link between y and x

Besides, we can also build another hexagon by considering two distinct argumentative relations linking a reason y to a conclusion x positively; see Fig. 5. Indeed, note the difference between (x, y) and $(\neg y \vee x, \top)$ which uncontroversially expresses that x entails y. Due to their structural differences they will not play the same role in LA: $(\neg y \vee x, \top)$ is stronger than (x, y), since $(\neg y \vee x, \top), (y, y) \vdash (x, y)$ (note that (y, y), i.e."y is a reason for concluding y" always holds).

Before concluding this short paper, let us suggest what kinds of attacks may exist between arguments in this setting. Let us take an example (adapted from [1]). Let us consider an argument such has "Mary little worked $(= y)$, she will fail her exam $(= x)$" will be represented by $(y, \top) \wedge (x, y)$. It may be attacked in different ways:

- by adding $(\neg y, \top)$ (No "Mary worked a lot");
- by adding $\neg(x, y)$ ("working little is not a reason for failing one's exam");
- by adding $(\neg x, y)$ (not very realistic here, although one might say "working little is a reason for not failing the exam (since one is not tired)");
- by adding $(\neg x, y \wedge z)$ ("Mary is gifted").

Note that the handling of this latter attack would require a nonmontonic inference mechanism, which may be encoded in a way taking lesson from the possibilistic logic approach [2], here based on the (partial) ordering defined by the propositional entailment: (x, y) should be drown if it exists a reason y', more specific than y, for concluding x' in a way opposite to x, i.e. we have both (x, y) and (x', y'), with x and x' inconsistent, $y' \vdash y$ (and $y \nvdash y'$).

4 Concluding remarks

Starting from linguistics-based evidence about argumentative statements, we have outlined the description of a two-layer logic, LA, for handling arguments, in conformity with a rich structure of oppositions which has been laid bare in terms of squares and hexagons. LA parallels generalized possibilistic logic. A more formal study of LA is the next step, as well as a comparison with other formal approach to argumentation. More lessons have to be taken from generalized possibilistic logic for handling the strength of arguments in a weighted extension, or for taking advantage of its relation to logic programming [7].

References

1. D. Apothéloz. Esquisse d'un catalogue des formes de la contre-argumentation. Travaux du Centre de Recherches Sémiologiques, 57, 69–86,1989.
2. S. Benferhat, D. Dubois, H. Prade. Practical handling of exception-tainted rules and independence information in possibilistic logic. Applied Intelligence, 9, 101–127, 1998.
3. J.-Y. Béziau. The power of the hexagon. Logica Universalis, 2012.
4. D. Dubois, H. Prade. Possibilistic logic: a retrospective and prospective view. Fuzzy Sets and Systems, 144, 3–23, 2004.
5. D. Dubois, H. Prade. Generalized possibilistic logic. Proc. 5th Inter. Conf. on Scalable Uncertainty Management (SUM'11), (S. Benferhat, J. Grant, eds.), Dayton, Oh., Oct.10-13, Springer, LNCS 6929, 428–432, 2011.
6. D. Dubois, H. Prade. From Blanchés hexagonal organization of concepts to formal concept analysis and possibility theory. Logica Universalis, 6, (1), 149–169, 2012.
7. D. Dubois, H. Prade, S. Schockaert. Règles et métarègles en théorie des possibilités. De la logique possibiliste à la programmation par ensembles-réponses. Revue d'Intelligence Artificielle, 26 (1-2), 63–83, 2012.
8. C. Lafage, J. Lang, R. Sabbadin. A logic of supporters. In: Information, Uncertainty and Fusion. (B. Bouchon, R. R. Yager, L. A. Zadeh, eds.), Kluwer, 381–392, 1999.
9. A. Moretti. Argumentation theory and the geometry of opposition (abstract). 7th Conf. of the Inter. Soc. for the Study of Argumentation (ISSA'10), 2010.
10. T. Parsons. The traditional square of opposition. The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), E. N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/square/.
11. G. Quiroz and D. Apothéloz and P. Brandt. How counter-argumentation works. In: Argumentation Illuminated, (F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds), International Centre for the Study of Argumentation (SICSAT), Amsterdam,172–177, 1992.
12. C. Salavastru. Logique, Argumentation, Interprétation. L'Harmattan, Paris, 2007.