ARGUMENT MOLECULES:

A FUNCTIONAL REPRESENTATION OF ARGUMENT STRUCTURE

Lawrence Birnbaum

Yale University
Department of Computer Science
New Haven, Connecticut

Understanding an utterance in an argument crucially requires determining the evidential relations it bears to prior and subsequent propositions in the argument (Birnbaum *et al.*, 1980; Cohen, 1981). The memory representation of an argument should, accordingly, indicate which propositions a given proposition counts as evidence for (a *support* relation) or against (an *attack* relation), and which propositions support or attack it in turn. The representation of an argument can thus be viewed as a network of propositions connected by support or attack relations (an *argument graph*). Although this sort of representation can be motivated simply by the need to represent the *content* of an argument, it seems natural to ask whether such argument graphs might further possess any useful *structural* properties, abstracted from the specific propositions they relate.

Identifying "useful" structural properties depends, of course, on having some notion of what uses they might have. In a process model of argumentation, a structural property of the representation of an argument is useful if (and only if) it serves some functional role in understanding or generating utterances in the argument. For example, it seems likely that some structural features of the argument would play a role in determining how an utterance relates, via support or attack links, to other propositions in the argument (Flowers *et al.*, 1982; Reichman, 1981). That is, a representation of argument structure should provide expectations as to which prior propositions, if any, a given input would be likely to attack, support, or be supported by. Such information would potentially be useful in reducing the number of prior propositions with which an input must be compared to determine whether or not an evidential relation holds.

A representation of argument structure should also help identify those propositions in an argument which are likely candidates for an arguer himself to attack or support, and thus play a role in planning a rebuttal. Explicit planning of this sort is not as central as might at first be supposed, however, because engaging in an argument is an *opportunistic* process (McGuire *et al.*, 1981), in which a good rebuttal to an input can often be discovered as a side-effect of the memory processing that is required simply in order to understand that input. One implication of opportunistic processing is that if a direct attack on an input is not found at understanding time, it will most likely be difficult to find one. In these cases, better chances for producing a good rebuttal probably lie elsewhere. Thus, a key function of explicit rebuttal planning, and hence of any structural considerations that enter into such planning, must be to focus attention on other points of possible contention in the argument when no direct attack on the input is discovered opportunistically.

One way to investigate argument structure is to attempt to identify commonly occurring patterns of support and attack relations that encompass several propositions. Because of their relatively fixed structure, these *argument molecules* (Flowers *et al.*, 1982) can be used to specify which propositions, among those contained in the molecule, are worth trying to attack or support. They can thus be used both to help plan rebuttals, and to generate expectations about an opponent's possible rebuttals. Two kinds of molecules have been identified thus far. The rest of the paper will illustrate the use of one of them in understanding and rebutting. (The other is described in Flowers *et al.*, 1982.)

A *stand-off* is an argument molecule involving a kind of *reductio ad absurdum* argument to attack an opponent's use of a plausible inference rule in support of one of his points. The attack is effected by showing that the opponent's reasoning can also be used to support some proposition that he cannot accept. For example, consider the following exchange in a mock argument between an Arab and an Israeli over Middle East affairs:

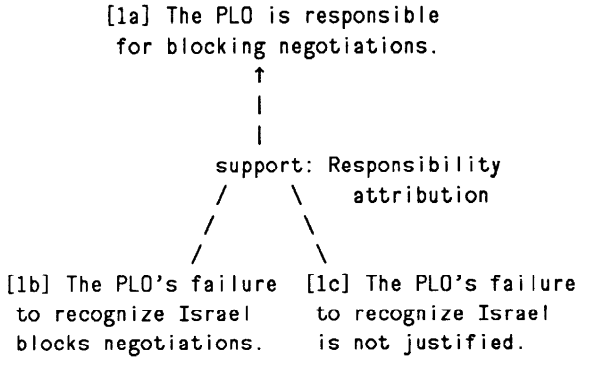
> [1] Israeli: Israel can't negotiate with the PLO because they don't even recognize Israel's right to exist.

[2] Arab: Israel doesn't recognize the PLO either.

The Israeli's utterance [1] is clearly an attack of some kind, but what exactly is being attacked? The explicit content of the utterance is that the PLO's failure to recognize Israel is blocking negotiations, and this in turn is based on the assumption that recognition of some sort is always a prerequisite for negotiations. However, this explicit content by itself does not attack anything. The force of the utterance stems instead from an implicit appeal to the notion of responsibility, using an interpretation rule which can informally be stated as follows:

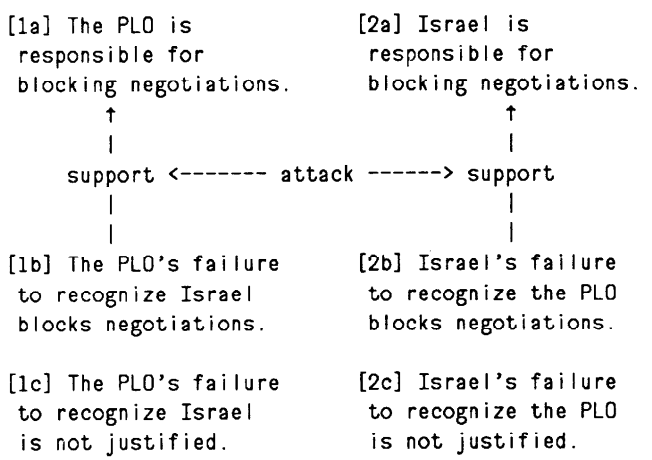
> *Responsibility attribution*: If an actor performs some action (or refrains from performing some action) which causes (prevents) some state of affairs, then that actor is responsible for causing (blocking) that state of affairs, unless its action (failure to act) was justified by some previous state of affairs.

The point of the Israeli's utterance, therefore, is that the PLO is responsible for blocking negotiations -- which is a *personal attack* (Flowers *et al.*, 1982) on the Arab position. The explicit content of the utterance counts as evidence in support of this claim, by appeal to the responsibility attribution rule. (Carbonell, 1981, has investigated the role of other interpretation rules of this sort in ideology-based reasoning about political events.) Thus, the proper representation of utterance [1] consists of the following fragment of argument graph:



Proposition [1c] represents the "unless" clause of the responsibility attribution rule. (This analysis of support relations is similar to that proposed by Toulmin, 1958.)

Now, the point of the Arab's response [2] is that, by using the same reasoning that the Israeli invokes in [1], Israel as well can be held responsible for blocking negotiations. Presumably, an Israeli would find this proposition as unacceptable as an Arab would find the Israeli's original claim. Thus, if the Arab is successful, neither disputant can use this line of reasoning without hurting his own position -- which is why this molecule is called a stand-off. The argument graph for the exchange has the following structure:



The utility of this structure in planning a rebuttal becomes clear when we examine the Israeli's options in response to [2]. One possibility is to attack the basis of the Arab's argument, in this case [2b] or [2c], for example by arguing that Israel does in fact recognize the PLO. However, if such an attack were possible, it would presumably be discovered opportunistically at understanding time. If so, there is no need to plan a response; if not, then the possibility of attacking [2b] or [2c] seems remote, and so should not be attempted. An attack on the Arab's claim [2a], that Israel is responsible for blocking negotiations, would be pointless, because the thrust of the Arab's argument is not that [2a] must be accepted, but that [2a] must be accepted if [1a] is. An attack on the support relation between between [2b] and [2a] would entail an attack on the responsibility attribution rule, which would be fatal to the Israeli's own claim that [1b] supports [1a]. And finally, there is clearly no point in trying to support [1b] or [1c], e.g., by proving that the PLO does indeed refuse to recognize Israel, because the Arab did not dispute it. Thus, within the scope of this stand-off molecule, the Israeli actually has only *one* option if no rebuttal arises opportunistically: he can attempt to re-support his claim [1a], that the PLO is responsible for blocking negotiations, using different evidence. Failing that, he can of course either concede the point or change the subject.

The stand-off molecule can play a similar role in the

process of understanding utterances in an argument. Consider the following continuation of the above exchange:

> [3] Israeli: But the PLO is just a bunch of terrorists.

This response can be accounted for by opportunism, because in the course of trying to understand [2], the Israeli must relate [2c], the claim that Israel's failure to recognize the PLO is not justified, to memory. In doing so, he will discover his belief that Israel's failure to recognize the PLO *is* justified, by virtue of the fact that the PLO is a terrorist organization. Because this response is simply an attack on the Arab's previous utterance, a possibility which should always be expected, it might seem that the stand-off molecule has no role to play in determining how it relates to the argument graph. This is not the case, however, because the representation of the Arab's utterance [2], when its implications are understood, consists of three distinct propositions, plus a support link. From the above analysis of the Israeli's options for rebuttal, it immediately follows that [3] cannot attack [2a] or the support link: only [2b] and [2c] are possible targets.

In order to employ the constraints associated with argument molecules to understand or produce responses, there remains the problem of recognizing such structures when they arise in an argument. If it turns out that there is a relatively small number of molecules, then the most straightforward method would simply be to attempt to match the templates of all of them against the most recent portion of the argument graph. A more interesting approach would involve identifying features of an utterance or an exchange which might typically signal some molecule. For example, an utterance of the form "But X does Y also" is probably a good indication of a stand-off molecule.

The examples presented here show that an argument molecule can be extremely useful in focusing attention on relevant propositions in the argument graph, both for understanding and rebutting. In essence, a molecule packages knowledge about the logical structure of an argument fragment in a way that makes explicit which potential responses would have some logical force, and which would not. We can conclude that arguments do indeed possess useful structural properties, abstracted from the specific propositions they encompass.

Acknowledgments: Many of the ideas in this paper grew out of discussions with Rod McGuire. Mark Burstein and Steven Salzberg provided useful comments on an earlier draft. This work was supported in part by the Defense Advanced Research Projects Agency, monitored by the Office of Naval Research under contract N00014-75-C-1111, and in part by the National Science Foundation under grant IST7918463.

References

Birnbaum, L., Flowers, M., and McGuire, R. 1980. Towards an AI model of argumentation. In *Proceedings of the First AAAI Conference*, Stanford, CA, pp. 313-315.

Carbonell, J. 1981. *Subjective Understanding: Computer Models of Belief Systems.* UMI Research Press, Ann Arbor, MI.

Cohen, R. 1981. Investigation of processing strategies for the structural analysis of arguments. In *Proceedings of the 19th ACL Conference*, Stanford, CA, pp. 71-75.

Flowers, M., McGuire, R., and Birnbaum, L. 1982. Adversary arguments and the logic of personal attacks. In W. Lehnert and M. Ringle, eds., *Strategies for Natural Language Processing*, Lawrence Erlbaum, Hillsdale, NJ, pp. 275-294.

McGuire, R., Birnbaum, L., and Flowers, M. 1981. Opportunistic processing in arguments. In *Proceedings of the Seventh IJCAI*, Vancouver, B.C., pp. 58-60.

Reichman, R. 1981. Modeling informal debates. In *Proceedings of the Seventh IJCAI*, Vancouver, B.C., pp. 19-24.

Toulmin, S. 1958. *The Uses of Argument.* Cambridge University Press, Cambridge, England.