

On Being Contradictory

Margot Flowers

Yale University
Department of Computer Science
PO Box 2158 (YS)
New Haven, CT 06520

Abstract

This paper discusses recognizing and producing contradictions. After illustrating the phenomena of contradiction, the paper presents conceptual classes of contradiction and gives an overview of the how they can be recognized. The next part discusses the construction of contradictions, in particular with respect to contradicting historic events. The object of this paper is to examine the computational logic of contradictions, using contradictions as an example of how reasoning processes can and must exploit semantic knowledge and episodic memory, and to illustrate the kind of metaknowledge needed to use certain reasoning devices correctly and effectively.

1. Introduction

How do we know when statements agree with one another and when they do not? Consider:

The Arabs started the 67 war. [a]

The Israelis started the 67 war. [b]

Most people would say that these two statements contradict one another, and the analysis they might provide is that [a] and [b] are similar statements which differ only by the actors. According to this analysis, then, [c] and [d] below should also contradict one another:

The Arabs imported arms. [c]

The Israelis imported arms. [d]

Obviously something is wrong with this analysis. The process of recognizing contradictions involves much more than simple structural matching.

How can the differences between [a]/[b] and [c]/[d] be explained? It is clear that any technique of determining what is a contradiction and what is not must hinge on *conceptual meaning*. Any technique based on structural form alone would fail, as illustrated above. The question to be explored here can then be rephrased as:

What conceptual features determine contradiction or agreement relationships between concepts?

This analysis of contradictions began in the context of ABDUL/ILANA, which models the participant in an argument about the Mid-East [Flowers, McGuire & Birnbaum, 1982, McGuire, Birnbaum & Flowers, 1981, Birnbaum, Flowers & McGuire, 1980], and the techniques described here are cur-

rently being incorporated into HARRY [Flowers, 1982], a program that models how people reason about historic events.

The approach taken here is that two concepts are *contradictory* when people perceive that only one of them can be possible or true at the same time. Thus, this analysis is not concerned with logical contradictions, but rather contradictions as people recognize them and use them.

An important aspect of the contradiction relation is that it is dynamic and context-dependent. The background knowledge that people bring to the process of conceptual analysis affects whether two concepts are interpreted as being contradictory or not. This is responsible for the fact that, among other things, additional information can cause seeming contradictions to be reanalyzed so that they are no longer contradictory. Consider:

John walked out on Mary yesterday.

That can't be true, I just saw her today and she was fine.

So far these sentences seem to contradict one another, but with the added information:

Oh, that is because they made up this morning.

what had seemed to be contradictory above, is no longer so.

Contradiction is only one of many possible relationships that can exist between two concepts. Thus, the detection of contradiction is really only a part of a more general analysis process. The reason for addressing contradiction here explicitly is to identify and enumerate what must be included in the process of conceptual analysis, and just as importantly the process conceptual generation, to account for the phenomena of contradiction. Thus, two major topics of this paper are concerned with contradiction recognition and contradiction generation.

2. The Logic of Contradictions

Often in explanations and arguments, for example, contradictory concepts are juxtaposed. For example, consider this fragment from ABDUL/ILANA:

Israeli: The Arabs [started the 67 war], by the blockade. [e]

Arab: But Israel fired first. [f]

Here, the Arab contradicts the assertion that the Arabs started the 1967 war by introducing the fact that Israel attacked first. So recognizing that there is a contradiction and what is being contradicted is a necessary part of the understanding process. How are any two concepts analyzed to determine this relationship?

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 contract IST7918463.

In this paper I present three classes of contradictions which can be identified according to how the concept is being contradicted. These classes identify how the analysis of potential contradictions must proceed by specifying what the significant parts of the concepts are and what related semantic knowledge is significant. These three classes are presented next.

2.1. Negations

The first class of contradiction is *contradiction by negation*. For example:

The Arabs started the 67 war.

The Arabs did not start the 67 war.

This kind of contradiction is commonly identified in textbooks on logic and rhetoric. The primary source of the contradiction relationship in this case is that the action of one is negated by the modality of the action of the other, with rest of the two concepts matching. In some cases, negation contradictions can sometimes be identified on structural principles alone, without the application of semantic knowledge. Consider:

The British srewolfed the tograms.

The British did not srewolf the tograms.

Although the meaning of these sentences is not clear, it is still obvious that they contradict one another.

When the action negation is obvious at a structural level, contradiction recognition is not very difficult or interesting. But more difficult cases occur, for example:

John payed attention to Mary.

John ignored Mary.

Here too, one sentence negates the other, but this time at a conceptual level. That is, the action "not paying attention" is the same as "ignoring". Thus, the recognition of negativity cannot rely upon the analysis of structural properties alone. Such semantic information is even more important in the other two classes of contradiction.

2.2. Direct Contradiction

The second class of contradiction is *direct contradiction*. For example:

The Iraqis caused the Persian Gulf war. [g]

The Iranians caused the Persian Gulf war. [h]

Why are these two sentences contradictory when others which resemble them are not? What is different about how they are analyzed that affects how they are related to one another?

Here the source of contradiction is lies in the interaction of certain conceptual components rather than in action negation. The key to determining this is based on general knowledge about wars and causation. In this case, we know that [g]/[h] contradict because, from our general knowledge about wars, we know that there can be only one primary instigator.

In general, determining if one concept directly contradicts another hinges on recognizing that they describe the same idea except for the presence of contradictory role fillers. This requires analyzing the semantics of the concepts involved to determine when dissimilar role fillers actually conflict. For example, the internal representation of [g] is:

(M-LEAD-TO
ANTE (DO ACTOR *IRAQ*)
CONSE E-PERSIAN-GULF-WAR)

which says that Iran did something that started the Persian Gulf War. Similarly, [h] is represented as:

(M-LEAD-TO
ANTE (DO ACTOR *IRAN*)
CONSE E-PERSIAN-GULF-WAR)

In this case, the only difference between these concepts is the antecedent component of the two LEAD-TOs, specifically the ACTOR role.

The goal here is to model the processes and assumptions used by people, naive about causation, who analyze and reason about situations. Among unsophisticated arguers over historic events, wars and other events have primary instigators: single *responsible intentional cause-ers* (although perhaps multiple causes). Here, [g]/[h] above are identified as contradictions after examining the general semantic knowledge associated with M-WAR and seeing that it allows there to be only one cause-er. This means that both of the antecedents cannot be causes, thus the concepts directly contradict one another.

The importance of general knowledge is assessing direct contradictions is illustrated by contrasting the analysis above with:

The Iranians fought in the Persian Gulf war.

The Iraqis fought in the Persian Gulf war.

The analysis of these concepts also must refer to general knowledge about M-WAR as before, but in this case, it utilizes the knowledge about *participants* of wars. The only limitation on the number of participants in a war is that there must be more than one, thus these two sentences do not present a problem in interpreting them both to be valid at the same time.

2.3. Inferential Contradictions

There is a large body of contradictions that do not fall into either of the two classes above, such as [e]/[f] above, and [i]/[j] below:

Arab: Israel is trying to take over the Mid-East. [k]

Israeli: If that were true, then how come Israel didn't take Cairo in 73? [j]

Here there is neither action negation nor contradictory role fillers. What then is the relationship between them? It is clear that there is some relationship that allows us to easily see [j] as contradicting [i], the problem is determining what this is.

The solution is to notice that, from [j], one can make the *inference* that:

Israel is not taking over the Middle East. [k]

Then, [k], which is supported by [j], negates [i]. In general, from an inferential contradiction, one can infer either a negation or a direct contradiction. Inferences such as these are one source of *support* and *attack* links used in the construction of argument graphs in ABDUL/ILANA. This then is an example of the third class: *contradiction by inference*, in which the inferred conceptualization establishes the link between two contradictory concepts.

Inferential contradictions are often the most effective types of contradictions. Negations and direct contradictions have no weight because they only state the fact of disagree-

ment, but provide no basis for that fact. Inferential contradictions on the other hand not only convey contention implicitly, but also explicitly provide a reason that contention is true.

Thus, if we consider the construction of contradictions rather than their recognition, it is clear that seeking inferential contradictions is one heuristic to use. This leads to the second part of this paper: overviewing how contradictions are produced.

3. Producing Contradictions

There are a many reasons one might want to produce contradictions, for example, in response to a strategic decision to attack some concept, as part of a process to confirm or decide to disbelieve something some idea, or to verify input. Thus, the production of contradictions is embedded in the context of some other task and often are never produced *explicitly*. For example, in arguments, ABDUL/ILANA takes advantage of the fact that the best attacks are contradictions that are noticed during the understanding process, rather than ones which are explicitly sought [McGuire, Birnbaum & Flowers, 1981].

The first two types of contradiction suggest two techniques to produce arbitrary contradictions. One way is to mechanically construct a concept's negation:

The West Bank belongs to Israel. [l]

The West Bank doesn't belong to Israel.

This, as pointed out above, requires minimal semantic knowledge. A second way is to replace some arbitrary critical role filler with something conflicting:

The West Bank belongs to Jordan. [m]

There are many problems with this approach. First of all, people are not usually called upon to produce some arbitrary contradiction to a given a disembodied concept. Secondly, the motivations of why a contradiction might be desired play a large part in what makes a good contradiction and what makes a bad one. For example, [m] is a good contradiction of [l] for a Jordanian to make, but a Palestinian would prefer instead to say:

The West Bank belongs to the Palestinian people.

Randomly changing one slot filler to another contradictory one does not reflect the *plans* and *goals* of the contradictor and is a fatuous thing to do.

The lack of episodic context is another problem with this suggestion of arbitrary contradiction construction. Without an episodic memory providing relevant information, which anchors potential contradictions to the real world, to contradict [l] one could just as easily say:

The West Bank belongs to Britain.

The West Bank belongs to Iran.

The West Bank belongs to Japan.

The West Bank belongs to Al Haig.

These are useless contradictions.

The ideal contradiction, for use in a naturalistic task like explanation or argumentation, obviously is one which has some validity in the world. Thus, the best contradictions are produced, perhaps as side effects, by *memory recall*, rather than the application of general semantic knowledge in the right way.

And since inferential contradictions are the most effective kinds of contradiction, inference classes provide useful heuristics for controlling memory search for contradictions. The key then is to determine what the useful kinds of inferential relations are. Here I will focus on making contradictions to historical events.

3.1. Contradicting Historic Events

A powerful way to contradict an assertion about an historic event is to use a *counterexample*. For example, [i]/[j] above, or:

Israel is not trying to take over the Middle East. [n]

Why then does Israel keep the West Bank? [o]

Thus, looking in memory for a counterexample is a good way to produce a contradiction. One way counterexamples can be found is to exploit previously noticed *failed expectations*, for example, a time Israel was expected to behave imperialistically but did not. The details of this process are discussed elsewhere [Schank, 1982, Flowers, McGuire & Birnbaum, 1982].

There are other sources of counterexamples, for example based in memory organization:

Israel isn't executing any terrorists. [p]

Yes they are, they killed three last week. [q]

The trouble is that seeming counterexamples do not always effectively produce a contradiction. For example:

Israel is importing arms. [r]

Israel did not import any arms last tuesday. [s]

Thus, "counterexamples" don't apply all the time. Memory processes have to know when they should consider producing contradictions and when they should not. How can this be determined?

The viability of producing a contradiction can be broken down into two conditions:

1. the *memory condition*: Is there present some fact which can serve as a contradiction?
2. the *semantic condition*: Is there a contradiction technique, such as counterexample, which is applicable to the concept to be contradicted?

Meeting the memory condition is of course entirely dependent upon what is in memory and the access techniques used to find it. The principles underlying the semantic condition however depend upon the contradiction candidate itself. What are these principles?

The counterexamples given so far illustrate one principle. The ones which "work" ([n]/[o] and [p]/[q]) have been cases of saying:

Someone is not doing an action.

Yes they are, here is an instance.

The ones which instead "fail" are cases of asserting:

Someone is doing an action.

No they aren't, here is a time they didn't.

They illustrate a variation of the well known fact that: in order to disprove a rule, all you need to is find one case that violates it; but to show it, it is not sufficient to illustrate one case that conforms to it, you have to show that every case conforms to it. This then identifies part of the semantic condition on the use of counterexamples:

Counterexamples are effective in contradicting an assertion that some action or state is not occurring.

However, this does not describe all of the aspects of the semantic condition. For example, there are cases in which counterexamples apply to contradicting *positive* assertions as well. Consider [i]/[j] above, or:

Israel is trying to take over the middle east.

Then why haven't they taken Lebanon.

This is a case of a counterexample being effective against a positive assertion. Clearly then, there are other principles involved in the semantic condition for making counterexamples.

The goal then is to identify what other class of principles which affect, among other things, representational criteria and applicable inference classes. HARRY uses four historical events classes that embody some of this information [Flowers, 1982]. Thus, historical event classes can be used to partially determine what kinds of contradictions are effective and should be sought when making a contradiction. They are:

atomic events -- historical events whose decomposition is not "interesting" at the current level of analysis, i.e. one is concerned with the initial attack by Israel in 1967, but not individual tank movements, etc.

episodic events -- sequences of causally/temporally linked events whose components are "interesting" and themselves are significant events, i.e. the 1967 war, which consists of attacks, battles, cease fires, and so forth.

stream events -- events which also consists of many sub-events but which are repeated instances of the same prototype, i.e. i.e. Israel importing arms.

interpretative events -- events based in the analysis of actions rather than in physical actions, often involving goals, of indefinite duration, and not composed of sub-events, i.e. i.e. Israel taking over the Mid-East.

Because historical events partially determine how events are represented and organized in memory, as well as the kinds of inferences that can apply, they can be used by memory access processes to identify potentially applicable contradictions and what kinds of memory components to seek.

The problem of determining when counterexamples are applicable to positive events instead of negative events is solved by applying more constraints on the semantic condition:

Counterexamples are effective in contradicting an assertion that some action or state is occurring only if the assertion is an interpretative event.

According to this principle, [n]; a positive interpretative event, can be effectively contradicted by counterexamples like [o], but positive stream events, like [r], cannot.

Some of the contradiction techniques whose applicability can be predicted by the historical event classes are: exemplars, counterexamples, instances, components, and initiating or terminating events. As with the examples above, not all of them apply to all kinds of events. Thus, the value of using these event classes is that they are one way of determining what to seek and what not to seek in memory in order to produce good contradictions, as well as for memory search and verification in general.

4. Conclusions

Contradictions are very important to a variety of tasks involving reasoning skills. This paper has given an overview of the kind of methods being incorporated into the reasoning model HARRY to recognize and produce contradictions and some of the information involved in doing so. The observations are these:

- Analyzing or recognizing contradictions is very different from producing them.
- Contradiction processing occurs primarily at the conceptual level, rather than at the structural level, and is grounded in and directed by semantic knowledge.
- Contradiction processing relies upon episodic memory.
- Contradiction processing is an interaction between many sources of knowledge, including the goals of both the producer of the contradiction and by the understander of the contradiction.
- Contradiction processing is a part of other processes: the understanding and generation of conceptual input can involve analyzing and producing contradictions.

Contradictions are only one of many reasoning devices. The use of any reasoning device requires knowledge about the use of the device itself and how it interacts with particular domains for it to be employed correctly and effectively.

Acknowledgements

The author is grateful to Michael G. Dyer, who made many useful editorial comments to a previous draft of this paper. The author also wants to acknowledge Larry Birnbaum and Rod McGuire, the other members of the ABDUL/LANA project. Aspects of event classes have benefited from discussions with Gregg Collins, Ernie Davis, and Michael Wilk, as well as with Larry and Rod.

References

- Birnbaum, Lawrence, Margot Flowers, and Rod McGuire. Towards an AI Model of Argumentation. *Proceedings of the First AAAI Conference (Stanford)*, 1980.
- Flowers, Margot. *Reasoning and Memory: A Computer Model of Human Reasoning*. PhD thesis, Yale University, 1982.
- Flowers, Margot, Rod McGuire, and Larry Birnbaum. Adversary Arguments and the Logic of Personal Attacks. In Wendy Lehnert and Martin Ringle (Ed.), *Strategies for Natural Language Processing*, pp 275-293, Lawrence Earlbaum Associates, 1982.
- McGuire, Rod, Lawrence Birnbaum, and Margot Flowers. Opportunistic Processing in Arguments, pages 58-60. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, 1981.
- Schank, R. C. *Dynamic Memory: A Theory of Learning in Computers and People*. Cambridge University Press, 1982.