Intent to Deceive:
On Creating Deceptions

Gregory B. Taylor

Artificial Intelligence Project
Department of Information and Computer Science
University of California
Irvine, Ca. 92717

ABSTRACT

Counterplanning can be successfully used against most methods of resolving goal conflicts. However, if one's intentions are disguised by deception then an opposing actor will use incorrect counterplanning or possibly none at all. This paper describes two components in the creation of a deception, the deception type and the enablement type, the range of their possible values, and how the selection of each can be used to create different deceptions for the same situation.

1. INTRODUCTION

Much work has been directed at understanding how people interact in a planned and intentional manner to resolve their goal conflicts (Schank [1977], Wilensky [1978]). Consider the following example,

> [1] John and Mary, brother and sister, wanted to watch different tv programs at the same time. There was only one tv set and both knew the other wasn't going to give in. John threatened to hit Mary if she did not let him watch his program.

The possible plans for resolving John's conflict can be ordered based on their likelihood of success and their difficulty in execution. Mary will counterplan against John based on her knowledge of his plan to resolve the conflict between them (Carbonell [1979]). For example, the story might of ended,

> [1a] Mary told their mother that John had threatened her. The mother sent John to his room.

John's ignorance of Mary's possible counterplans resulted in his goal failure.

Deceptions are a class of plans where the intentions of the deceiver are purposefully not communicated thereby preventing successful counterplanning. For example, John could have deceived Mary by,

> [1b] John mentioned to Mary that he had seen several girls going to the theatre to see Robert Redford making a public appearance. Mary immediately left, leaving John to watch his program.

John led Mary to believe that his intentions were to give her a chance to meet Robert Redford, when in fact he simply wanted her out of the house.

This work was supported in part by the Naval Ocean Systems Center under contract N00123-81-C-1078.

2. HOW DID JOHN LIE?

Here we briefly raise the question, "HOW DID JOHN KNOW TO LIE?" As mentioned earlier, John's selection of a plan to resolve the conflict (to lie) is based on how likely that play is to succeed. Deceptions are most successful when the relationship between two people is seen as a benevolent from the perspective of the person to be deceived (Mary) and malice from the perspective of the deceiver (John). For example, does she trust him, does he dislike her, etc. Opportunistic aspects of deceptions also exist and can be recognized by characters.

In this section we introduce the main topic of this paper - the two components of a deception, the deception type and the enablement type, and show how they combine to create a deception such as the one in 1b.

2.1 Deception types

There are four major classes of deception types or d-types. Each class of d-types is used to either achieve the deceiver's goal or cause the deceived person plans to fail. The deception occurs when the deceived person is not aware of what is happening.

D-types in the first class select a precondition of the deceived person's plan to be negated. By "undoing" a precondition that is difficult to re-establish the deceived person's plan will fail. The original deception goal is reduced to negating this precondition. The negation of this precondition becomes the new deception goal. If the actual goal of the deception is to prevent some action on the part of the deceived person then this d-type is very useful. The d-types are described in the first person (I deceive you).

1. Undo preconditions for object within a plan. The new deception goal is to make you believe that the particular desired attributes for some required object in your plan no longer exist. For example, if I want you to leave the apple pie so that I can eat it, I might tell you that the pie is rotten; if John wants Mary not to watch television, he convinces her that the set is broken.

2. Undo delta goal preconditions (Schank [1977]). The new deception goal suggested here is to undo any one of the simple preconditions commonly found within the deceived person's plan such as control of an object, being at a location or knowledge of some simple fact.

Because delta goals appear in most plans, the conflict with the actual goal is difficult to notice. In 1b, John uses the undo delta goal d-type (PROX is selected). The new goal to be achieved is to undo Mary near the television, i.e. to make Mary leave the television area.

D-types in the second class of the four classes are useful when the deception goal is a simple action and incorporate it into some larger action (backwards from the previous class of d-types). The simple action is usually a delta goal, although not necessarily. It is often the result of reducing an original deception goal using a d-type from the first class of d-types. The d-types of the second class are described below (again in the first person).

3. Challenge. I identify a plan that contains you acting out the present deception goal as a small part or precondition. The plan should contain use of some boastful attribute, e.g. strength, quickness, singing, etc. I use reverse psychology in communicating the plan. For example if a mother wants her son to empty the trash, she says, "I bet you're not strong enough to empty the trash with one arm."

4. Demonstration. Similar to challenge but I communicate the plan in a straight forward manner - no reverse psychology.

5. Instantiate common context. We call a commonly done activity a context. I locate one of your contexts that has the present deception goal as one of its preconditions or steps in execution. I communicate or instantiate the goal or intention of the context to you and play out the context UNTIL the deception goal is reached.

6. Posit better goal. Very similar to instantiate common context, except that the activity is not commonly done. Here, I must propose an explicit goal for you to pursue. It must be of higher importance to you than your present goal. Often the new goal is just an instantiation of the present goal with some parameter changed to effect the greater value.

The common feature of these d-types is that the plan selected should contain some boastful attribute that can be used to "emotionally push" the deceived person into enacting the plan.

Recall in 1b, John's new deception goal is to make Mary leave the television area. Here John uses the posit better goal d-type. The goal selected is the same as Mary's present goal - that of entertainment. However the restrictions on the exact construction of the goal are that the location of the entertainment be away from the television and that it be more "interesting" than the program that Mary was going to watch. From these descriptions, John creates the goal that Mary should go to the theatre to see Robert Redford. The deception is far from complete, he still he still must make her believe that Robert Redford is at the theatre. However, it is unlikely that she will see the connection between Robert Redford at the theatre and watching TV.

2.1.1 Structure of a deception

John's deception has used two d-types. We can represent the present structure of what we have analyzed.

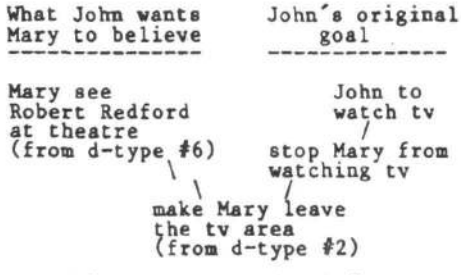

Figure 2-1: Structure of John's deception

The number of d-types used in a deception depends on how many are necessary before a "believable" deception is reached. That is, after John's first d-type, it is unlikely that Mary will leave the television. Making a "believable reason" for her to leave is accomplished by "stretching" the distance from the original deception goal to the final deception goal.

The length of the deception (number of d-types used) obscures John's original intentions/goals. Also, as Mary views his goals as "less conflicting" to hers, his d-types become more believable to her. She believes that John is no longer out to compete with her for the television, but instead he is her benefactor.

2.1.2 Other d-types

D-types in the third class of d-types distract the deceived person by communicating the existence of some very high priority goal. These d-types are similar to the posit better goal d-type except that the new goal presented is one of such high priority that the decision of which goal to pursue to not made, rather achieving the previous goal is postponed. These d-types are listed below, again in the first person.

7. Attract attention. I communicate the existence of a very high opportunistic goal that you can easily achieve such as money on the ground or seeing a beautiful girl walking down the street.

8. Crisis goal violation. I communicate the possible violation of one of your crisis goals such as maintain-health.

D-types in the final class deal with more complicated goal modification and interaction. Deceptions are introduced by convincing the deceived person of false goal relationships. For example,

[2] John was spraying the garden with pesticides. His wife who didn't like chemicals told him that she knew from biology that pesticides kill cats and that she was thinking about getting a cat. He stopped spraying.

Mary made John believe that spraying prevented a higher goal of her happiness (having a cat) from being achieved. Because of length limitations we are unable to individually describe these last d-types.

Another example of a deception in this class might be if we want to convince someone that a goal is undesirable or can't be achieved we might first show them "the entire" set of plans that can achieve their goal and then show how each plan is undesirable or can't be achieved. The deception in such a strategy could be in an incomplete breakdown of the possible plans, showing a good plan to be bad, or showing how an achievable plan will fail.

2.2 Enablement types

In example 1b, John's second d-type still leaves him with the deception goal of how to make Mary believe that Robert Redford is at the theatre. The d-types have been methods of altering the deception goal. Introducing the false fact occurs in the enablement type or e-type.

In general, e-types appeal to the emotions. Enablement types or e-types have two functions. First, appealing to the emotions of peer pressure or the feeling of "not wanting to miss anything" as in John's deception. Secondly, e-types cause the deceiver to believe different attributes about the deceiver. Either general attributes like feelings of trust and friendship which result in a decreased level of suspicion, or specific attributes regarding a particular piece of information. The e-types are listed below, again in the first person.

1. Others say or do. I make you believe some fact because I tell you that others believe it. For example, John said to Mary that "he had seen several other girls ...".

2. Complement you. I can say to you, "You are so nice ..." or complement you by comparing you favorably to your enemy. Also included are attributing your resent goal failures to your enemies.

3. Do favor. I can help you achieve a goal. I chose a goal where my assistance is necessary to achieve a goal that you are currently pursuing or have abandoned because of a plan failure. For example, "Would you like some help on your calculus problems?"

4. Knowledge and experience. I can convince you that I have some specific knowledge. For example, "I know how a tv works 'cause I took a class ... and this one is busted." When in fact it is simply unplugged.

5. Experience yielding specific attributes. I can "prove" to you that some past experience has happened to me by relating specific details of the experience. The inference that relates the desired attributes to the experience must be known to the person being deceived. For example, if I am a woman and know men only seriously date women who have dated before, I might carry a locket showing a picture of a man to whom I "would claim" I was engaged.

In John's deception of Mary, he has selected e-type #1. John tells Mary that "others are doing" the same goal (constructed using the previous d-types) he suggests for her. The selection of e-types and their "execution" is also aided by the previously used d-types. For example, the complement e-type is suggested by the demonstrate d-type as in the story below from Firman and Maltby [1918].

[3] A sparrow sitting on a log noticed a robin on a branch directly above him holding a worm in his mouth. The sparrow said to the robin, "Robin, you sing the most beautiful songs in all the forest - won't you sing for me now?". The robin always willing to show off opened his mouth to begin singing. The worm immediately fell out of the robin's mouth and dropped to the ground next to the sparrow. The sparrow quickly ate the worm and left.

The d-type demo of singing suggested the complement e-type along with what specifically to complement.

3. FUTURE WORK

A program is being written to use the first three classes of d-types with all the e-types to construct different deceptions. Each d-type and e-type will be a procedure that builds up the representation for the deception. The program will eventually be integrated into a simulation environment such as Tale-spin (Meehan [1976]).

3.1 A Deception Matrix

The current set of d-types and e-types can be used to create a deception matrix that results in every possible deception for a given situation. The information we hope to obtain from such an analysis includes how complete the d-types and e-types are in creating believable deceptions and which combinations tend to produce deceptions most nearly to those that humans produce.

3.2 Better d-types

Examples of deceptions that need to be understood in greater detail include third party deceptions. For example,

[4] John's love was not returned, she loved Bill instead. John wrote her a letter saying it was all over and signed it Bill.

Complex deceptions not using the d-types discussed exist and must be studied. For example, from our original television example, John's deception might have been,

[1c] John hid the tv guide from Mary and told her that her program had been cancelled. She left to go play.

How did he know to hide the television guide to make his deception work?

The forth class of d-types that we mentioned will probably always need more work. We have tried to categorize this class, however it still remains full of the most complex deceptions involving goal relationships.

3.3 Better e-types

Carbonell [1979] has used some basic personality traits to describe an individuals method of goal pursuit. But how do these same traits influence other goal interactions and believability? Also of interest are psychological theories of how people can be made to feel friendly towards others based only on common experiences and/or friends.

4. CONCLUSIONS

The role of identifying a set of d-types and e-types is not to necessarily produce every possible deception, only a large set of varying types of deceptions. To this end we have already succeeded; with 8 d-types and 5 e-types the number of possible deceptions for any situation is forty. However, in most cases only a few of these are "acceptable". Our future work will focus on increasing the possible deception plans in order to select the best deception plan.

REFERENCES

[1] Bruce, B. "Analysis of Interacting Plans as a Guide to the Understanding of Story Structure". Poetics 9 (1980) pp. 295-311.

[2] Carbonell, J. Subjective Understanding: Computer Models of Belief Systems. Ph.D. thesis. Yale Computer Science Department Research Report 150, 1979.

[3] Firman, S.G. and E.H. Maltby. 1918. The Winston readers: first reader. Philadelphia: Winston.

[4] Meehan, J. The Metanovel: Writing Stories by Computer. Yale University, Computer Science Department, Ph.D. thesis 1976.

[5] Schank, R. and Abelson R. Scripts, Plans, Goals and Understanding. Lawrence Erlbaum Associates, Hillsdale, N.J, 1977.

[6] Wilensky, Robert Understanding Goal Based Stories. Ph.D. thesis. Research Report 140, Yale University, Department of Computer Science, Yale, 1978.