Upholding the Maxim of Relevance during Patient-Centered Activities

Abigail S. Gertner and **Bonnie L. Webber**
Computer & Information Science
University of Pennsylvania
Philadelphia PA 19104-6389
agertner@linc.cis.upenn.edu
bonnie@central.cis.upenn.edu

John R. Clarke
Department of Surgery
Medical College of Pennsylvania
3300 Henry Avenue
Philadelphia PA 19129
jclarke@gradient.cis.upenn.edu *

Abstract

This paper addresses "kinds and focuses of relevance" that a language-generating clinical decision-support system should adhere to during activities in which a health care provider's attention is on his or her patient and not on a computer screen. During such "patient-centered" activities, utterances generated by a computer system intrude on patient management. They must be thus seen by HCPs as having immediate clinical relevance, or, like the continual ringing of ICU monitors, they will be ignored. This paper describes how plan recognition and plan evaluation can be used to achieve clinical relevance. The work is being done in the context of the TraumAID project, whose overall goal is to improve the delivery of quality trauma care during the *initial definitive phase* of patient management. Given an early pilot study that showed that physicians using Traum-AID disliked the continuous presentation of its entire management plan, we decided to explore how TraumAID could restrict commentary to only those situations in which a comment could make a clinically significant difference to patient management. We took advantage of the fact that actions that involve resources that need to be brought to the trauma bay or that can only be done elsewhere must be *ordered*. Since orders can be rescinded, comments pointing out problems with an order can potentially make a clinically significant difference to patient management. The contributions of this paper are (1) pointing out additional

constraints on language generation raised by the desire to convey information to listeners attending to something other than an computer terminal, and (2) pointing out some features of plan inference and evaluation raised by multiple goal planning in a complex domain.

1 Introduction

Ordinary use of Natural Language adheres to what Grice has called the "cooperative principle of conversation" (Grice, 1975). Four categories of "maxims" elaborate this principle. Under the category of *Relation*, Grice places the single maxim "Be relevant", noting that this terse formulation conceals a number of unsolved problems, including that of "what different kinds and focuses of relevance there may be".

This paper addresses "kinds and focuses of relevance" that a language-generating clinical decision-support system should adhere to during patient-centered activities – that is, activities in which a health care provider's (HCP) attention is on his or her patient and not on a computer screen.[1] Such activities include surgery, childbirth, and emergency medical care (including the responsibilities of Emergency Medical Technicians). During patient-centered activities, utterances generated by a computer system intrude on patient management. They must be seen by HCPs as having immediate clinical relevance, or, like the continual ringing of ICU monitors, they will be ignored. Utterances that would *not* make a clinically significant difference to patient management must be avoided. This paper describes how plan recognition and plan evaluation can be used to achieve this end.

This work is being done in the context of the TraumAID project (Clarke et al., 1993; Clarke et al., 1994; Rymon et al., 1993; Webber et al., 1992),

[1]In this sense, our emphasis differs from Sperber and Wilson (Sperber and Wilson, 1986), who are concerned with the problem of inferences that can be made under the assumption of relevance.

whose overall goal is to improve the delivery of quality trauma care during the *initial definitive phase* of patient management. Initial definitive management scopes a complex range of diagnostic and therapeutic procedures. Resolving often conflicting demands for managing multiple injuries requires global reasoning that can exceed the restricted locality of rule-based systems. The modular architecture of the current system, TraumAID 2.0, consists of two complementary processes: (1) a rule-based reasoner able to draw diagnostic conclusions and identify what goals each of them implies; and (2) a planner that takes the set of currently relevant goals and constructs a management plan – a partially-ordered sequence of actions – that is most appropriate at that point in time.

TraumAID 2.0 has been retrospectively validated against actual trauma management plans for 97 non-pregnant adult patients presenting consecutively to the Medical College of Pennsylvania (MCP) Trauma Center with penetrating injuries to the chest and/or abdomen. A blinded comparison of management plans was carried out by a panel of three experienced trauma surgeons. They preferred TraumAID 2.0's plans to actual management to a statistically significant extent (Clarke et al., 1993).

A decision support system will not be clinically viable without solving its interface problems. For us, this means (1) getting TraumAID the information it needs for its reasoning and planning, and (2) getting the results of TraumAID's reasoning to the members of the trauma team (in particular, to the chief surgical resident) during the course of care.

To get information into TraumAID, we have developed an electronic version of a standard *trauma flow sheet*, so that information can be entered by the scribe nurse during the trauma resuscitation.[2] The electronic trauma flow sheet is implemented in HyperCard, supported by a FoxPro database. Relevant information entered by the scribe nurse is automatically passed to TraumAID 2.0.

The focus of this paper is the problem of delivering the results of TraumAID's deliberations, given an early pilot study that showed that physicians using TraumAID disliked the continuous presentation of its entire management plan. (Currently, this is displayed on the nurse's monitor in a separate window.) The reasons for physician dissatisfaction appear to be that (1) much of the time, TraumAID's recom-

mendation coincided with the physicians' own plans, and (2) having the entire plan presented made it difficult for the physicians to determine what, if anything, they should focus on. To put it another way, presenting the entire plan violates Grice's Maxim of Relevance.

We decided to explore how TraumAID could restrict commentary to only those situations in which a comment would be clinically relevant in terms of its potential for making a clinically significant difference to patient management. We took advantage of the fact that actions that involve resources that need to be brought to the trauma bay or that can only be done elsewhere must be *ordered*. Since orders can be rescinded, comments pointing out problems with an order can potentially make a clinically significant difference to patient management. Our prototype interface for TraumAID produces commentary when an order differs significantly from what TraumAID would recommend. It also generates commentary when the physician fails to order something that TraumAID thinks should be done within the immediate time frame. The interface, TraumaTIQ (Gertner, 1994; Gertner, 1994b), uses plan inference and plan evaluation to recognize both such errors of commission and errors of omission. In this paper, we describe TraumAID's knowledge of management goals and actions that are used in plan inference and evaluation (Section 2), the plan recognition and evaluation strategies that allow TraumaTIQ to recognize clinically significant differences in patient management(Section 3), and the additional problems for plan recognition, plan evaluation and critique generation caused by a planner's ability to optimize plans through what Pollack has called "overloading intentions" (Pollack, 1991) (Section 4). We conclude with mention of our planned prospective evaluations of TraumAID and TraumaTIQ and a summary of our main points (Section 5).

2 The Representational Framework

In TraumAID, a plan is a set of *procedures*, each of which comprises one or more *actions*. Procedures are used to address diagnostic and therapeutic *management goals* motivated by what is currently known about the patient. A particular action can participate in more than one procedure, and thus can be used in addressing more than one goal. As noted earlier, TraumAID uses a *rule-based reasoner* to derive a set of relevant goals from the currently available knowledge about the patient and a *planner* to choose appropriate procedures to address those goals and schedules the actions they involve.

TraumAID's reasoner controls information acquisition using a conservative, *staged* strategy for diagnosis and treatment (Rymon, 1993): expensive, definitive tests are not included in a plan until they are justified by less costly tests or observations, and

---

[2]One member of the trauma team is a nurse who functions as a scribe, documenting all the findings, tests, and treatments in chronological order for the record. The multiple page trauma flow sheet has designated areas for specific information, such as demographics, mechanism of injury, physician response times, trauma score, Glasgow coma score, vital signs, location of wounds, results of primary assessment, intravenous therapy, diagnostic and therapeutic procedures, medications given, fluid intake and output, and disposition.
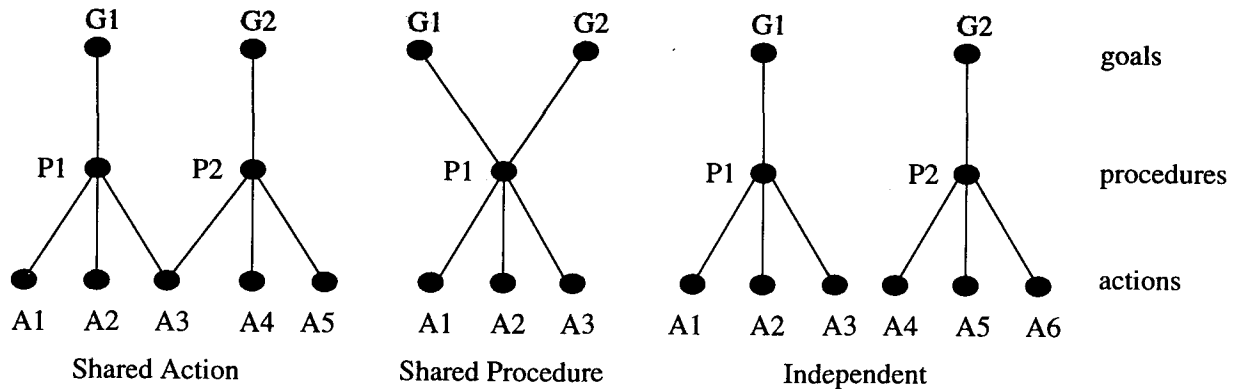
Figure 1: Three possible multiple goal-procedure-action configurations

definitive treatment is not recommended without the results of sufficient evidence from diagnostic tests.[3] These strategies are reflected in the knowledge base by the occurrence of related management goals, such as a goal to diagnose hematuria (blood in the urine), which if present, triggers a goal to diagnose bladder injury, which in turn can lead to a goal to treat bladder injury. Goals that do not participate together in a coherent strategy may still be connected by test results. For example, the goal of finding a bullet in the mediastinum leads to doing a lateral chest x-ray. While this might also reveal a fractured sternum, there is no strategic relationship between the goal of finding a bullet in the mediastinum and the goal of treating a fractured sternum.

Once a set of relevant goals has been determined, the planner's choice of how address each goal is based on both local and global considerations. Goals and procedures are linked via a *goal-procedure mapping* which lists alternative procedures for addressing each goal. The procedures in a mapping are listed in order of preference so that, all else being equal, the first procedure will be chosen to address the goal. Less preferred procedures may be selected if they would result in a globally more optimal (less costly) plan. Figure 1 shows three possible configurations for a portion of a plan addressing two goals. (Section 4 discusses the consequences for plan recognition, evaluation and critiquing, of the ability to create plans in which actions and/or procedures are shared between goals.)

Actions are scheduled according to both logistical and clinical considerations. Logistical considerations mean that patients are only moved in one direction through the Trauma Center – from the emergency center, optionally to the radiology suite, then optionally to the operating suite, and finally to the

trauma unit. Since actions may have constraints on where they can be performed, they are scheduled so as to avoid transferring the patient back to a place he has already been.

Clinical considerations have to do with the *urgency* and *priority* of each action, which it inherits from the goals it is being used to address. The urgency can be either *catastrophic*, *unstable*, or *stable*, representing the patient's condition and thus the amount of time available in which to address that goal. Catastrophic goals must be addressed immediately. Unstable goals must be addressed before stable goals. Priorities represent standard practices of trauma care: if there are no differences in urgency, problems involving the airway are addressed before those involving breathing, which are addressed before those involving circulation, etc. (the "ABCs of trauma care").

3 Interpreting and Reacting to Orders

We have claimed that in patient-centered activities, the effectiveness of a "by-stander" decision-support system depends on focusing its clinical role on the immediate needs of physicians rather than intervening with information that, while correct and apropos, would not make a significant difference to patient management. In order to achieve this, our approach is largely *reactive* rather than *proactive*. As noted in Section 1, instead of presenting TraumAID's recommended management plan, TraumaTIQ waits for the physician's orders, which are treated as intentions to perform those actions. TraumaTIQ interprets and evaluates these intentions, and generates a comment if it would be relevant to do so. TraumaTIQ also interprets the lack of an ordered action within an appropriate time frame as indicating the physician's lack of intention to perform that action and evaluates it as such. This approach is a form of critiquing, early examples of which are (Langlotz and Shortliffe, 1983; Miller, 1986).

---

[3]The one case where this is not true is when a patient comes in near death, with catastrophic chest wounds. Surgery is recommended immediately, without attempting to diagnose what specific injuries may have been sustained.

TraumaTIQ's critiquing process is triggered whenever new information is entered by the scribe nurse and delivered to TraumAID. This information can be in the form of (1) bedside findings, (2) diagnostic test results (indicating both that a diagnostic action has been performed and what the results of that action were), (3) therapeutic actions performed, or (4) diagnostic or therapeutic actions ordered by the physician. TraumaTIQ interprets the physician's orders in a goal-directed manner, using TraumAID's representation of goals, procedures and actions, so that the critique can address the likely reasons underlying any discrepancies and can suggest alternative means of addressing a particular goal. Figure 2 shows the architecture of TraumaTIQ, comprising *plan recognition, plan evaluation,* and *critique generation.*

3.1 Plan Recognition

In order to judge when it is relevant to comment on a physicians orders or lack thereof, it is useful to consider not only *what* actions have been ordered, but also *why.* For example, this can allow comments to be withheld if an ordered action that differs from TraumAID's choice would nevertheless satisfy pertinent management goals. TraumaTIQ's plan recognition component aims to infer the underlying goal structure motivating the physician's orders for actions to be performed. Note that this is an example of keyhole recognition, rather than intended recognition (Kautz, 1990). The physician does not *intend* to provide enough information to allow others to infer his plans. Like other members of the trauma team, TraumaTIQ must infer the plan incrementally based on what is known about the patient and what has been ordered so far.

Trauma management poses particular problems for plan recognition. First, there is no fully preset sequence in which actions are ordered and done. Secondly, multiple diagnostic and therapeutic goals may simultaneously be active, and a single action can often be used to address several related or unrelated goals. In this domain, it would thus be incorrect to simplify search by minimizing the number of top-level goals in the inferred plan, as Kautz and others have proposed (Kautz, 1990). Third, a plan recognition strategy cannot assume that the physician's plan is correct: management of trauma patients must often be done late at night and under time pressure, conditions that can lead to less than optimal decision-making even in trained professionals. Since the number of possible incorrect plans is too large to encode *a priori,* we must make certain assumptions about physicians' plans:

- The head of a trauma team will have expert or near-expert knowledge of trauma, and will usually develop plans that are similar to TraumAID's. Thus, if an action has been ordered that is also in TraumAID's plan, TraumaTIQ

assumes that it is being done for the same reason(s). We call this giving the physician the "benefit of the doubt."

- The physician is more likely to have appropriate goals but be addressing them in a sub-optimal way, than to be pursuing the wrong goals altogether.

- While TraumAID follows a conservative strategy for pursuing diagnosis and treatment from observations, physicians may proceed more rapidly, pursuing a goal that may be involved in a current strategy but for which TraumAID does not yet have enough evidence to conclude its relevance. An understanding of the strategic relationships between goals should help to recognize examples of this difference.

Reflecting these assumptions, the plan recognition algorithm works as follows:

1. When an action, α, is ordered by the physician, check whether α is currently a part of TraumAID's recommended plan as a means of satisfying all or part of goal γ, or all or part of each member of a set of goals Γ.

2. If so, add γ or Γ to the representation of the physician's plan.

3. If α is *not* currently in TraumAID's plan, determine whether there is a relevant goal that α might address:

   (a) If any of the goals that might lead to doing α are present in TraumAID's current set of active goals, assume that α is being done to address that goal or goals.

   (b) In the case that there is no relevant goal to explain why the physician is ordering α, check whether any of the possible goals motivating α are part of a currently active diagnostic strategy (cf. Section 2).

   (c) If no relevant goal or strategy is found, leave the goal unspecified and add the intention to do α to the representation of the physician's plan with no goal attached. There is one exception to this rule:

   (d) If the system only knows of one possible goal that would lead to performing α, TraumaTIQ assumes that α is being done to address that goal, even though it does not consider the goal to be relevant.

3.2 Plan Evaluation

After attempting to infer the goal(s) underlying the physician's actions, TraumaTIQ attempts to identify whether commentary is warranted on any aspects of his plan. Plan evaluation begins by comparing the plan attributed to the physician with the plan produced by TraumAID, looking for four types of discrepancies:
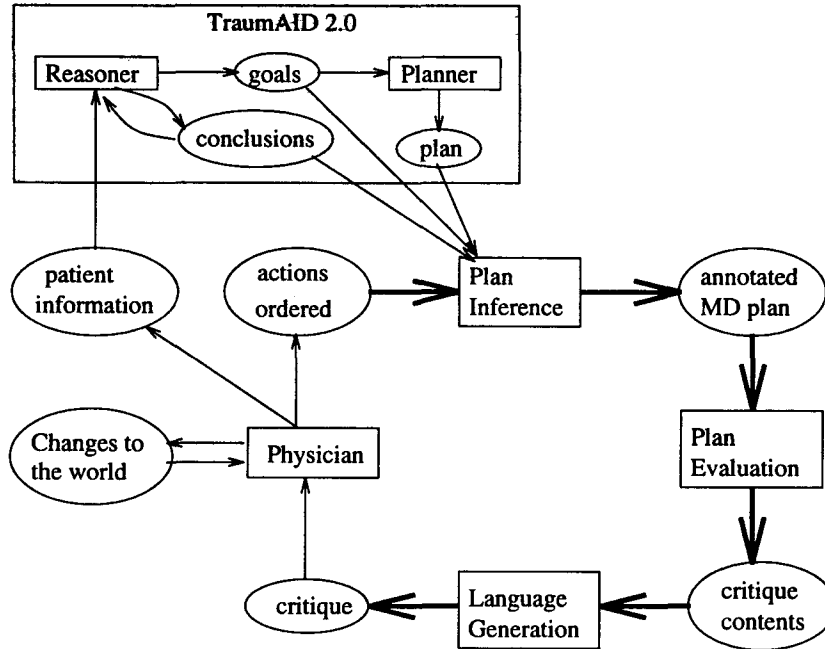
Figure 2: The TraumaTIQ module

- *Omission*: A goal that TraumAID considers relevant is not being addressed by the physician in a timely manner. This can be further analyzed as to whether (1) the goal is not being addressed at all, or (2) the goal is only being partially addressed – some but not all the actions in the procedure addressing the goal have been ordered.

- *Commission*: An action is present in the physician's plan that *does not* address a relevant goal. If a unique goal can be inferred to explain this discrepancy, that goal can be further categorized as to whether (1) it is unwarranted, (2) it is not fully proven, or (3) it has already been addressed.

- *Procedure choice*: A relevant goal is being addressed, but not using the procedure preferred by TraumAID.

- *Scheduling*: Actions are not being done in the order recommended by Traumaid, e.g., satisfying urgent goals before non-urgent ones.

TraumaTIQ upholds the Maxim of Relevance by attempting to avoid comments that may later prove irrelevant. Errors of omission are not critiqued until a sufficient period of time has elapsed during which the physician might order the action. The amount of time allowed depends on the urgency of the goal that the action is intended to address – the rule of thumb TraumaTIQ uses is that a comment should be produced after approximately 10% of the time period has passed that is available to address the

goal without significant consequences. Furthermore, comments are only made with respect to goals that cannot be made irrelevant by actions scheduled to be done before those goals are addressed.

Discrepancies of all types are then evaluated in terms of their *potential clinical significance*. Currently this is only a rough estimate based on approximate cost and whether or not a procedure is invasive. For example, an unnecessary chest x-ray is considered insignificant while an unnecessary (and invasive) laparotomy is considered clinically significant and worth drawing to the physician's attention. We are currently in the process of developing more objective criteria for classifying errors according to their potential impact on the outcome of the case. Each discrepancy will be classified as either: (1) *tolerable*, probably harmless, (2) *non-critical*, but potentially harmful, or (3) *critical*, potentially fatal. Anything in the second or third category will be considered significant enough to be reported in the critique, while tolerable errors will not be mentioned.

The output of plan evaluation is a set of *communicative goals* containing (1) a *propositional content* (PC) indicating the type of discrepancy and the particular TraumAID concepts involved, and (2) an *illocutionary force* (IF), such as URGE, INFORM or REMIND, indicating *how* the information should be realized linguistically. For example, the goal:

(SUGGEST
(PROCEDURE-CHOICE   GET-X-RAY-LAT-ABD
GET-CT-SCAN-ABD

129

RO-Compound-Fracture-Lumbar-
Vertebra))

would be produced in a situation in which the physician has ordered a CT-scan of the abdomen, which TraumaTIQ has inferred is intended to address the goal of diagnosing a compound fracture of the lumbar vertebra, which TraumAID has instead chosen to address with a considerably less costly, less time-consuming lateral abdominal X-ray. The illocutionary force of SUGGEST indicates that this is a non-critical error. Had it been evaluated as a critical error, the illocutionary force would have been URGE.

3.3 Critique Generation

Critique generation serves to organize TraumaTIQ's communicative goals according to the management goals they address and to translate them into Natural Language utterances. This is currently the least sophisticated part of the process: each IF-PC pair indexes a sentential template with syntactically marked slots to be filled in with the appropriate phrasal translation of a TraumAID concept. For example, the template indexed by (SUGGEST, PROCEDURE CHOICE) is:

> "TraumAID suggests (ARG1 GERUN-DIVE) rather than (ARG2 GERUN-DIVE). The former is preferred for (ARG3 GERUNDIVE)."

The first and second slots are filled in with gerundive phrases corresponding to GET-X-RAY-LAT-ABD and GET-CT-SCAN-ABD, while the third slot is filled in with an untensed verb phrase corresponding to RO-Compound-Fracture-Lumbar-Vertebra, resulting in the sentence:

> "TraumAID suggests getting a lateral X-Ray of the abdomen rather than getting a CT-scan of the abdomen. The former is preferred for checking for fracture of the lumbar vertebrae."

Critiques can be delivered directly to the trauma bay through synthesized speech or to the scribe nurse for subsequent delivery to the trauma team.

4 Action Overloading

As mentioned earlier, it is possible for a single action to participate in addressing more than one management goal. This strategy of doubling-up the use of plan elements is called *intention overloading* by Pollack (Pollack, 1991), who argues that it can simplify the process of computing optimal plans. It can, however, complicate the process of *recognizing* and *evaluating* them. Our strategy in TraumaTIQ is to assume that overloading occurs in the physician's plan whenever it would be beneficial. In other words, if an action ordered by the physician can be used to address two or more relevant goals, it is assumed that

that is what the physician intends. This assumption makes sense because as long as both goals are being addressed, there is no need to produce a critique, even if the physician did not have both goals in mind.

If, on the other hand, there is a possibility for overloading that the physician does not exploit, it may be clinically relevant to mention it. For example, consider a patient with a gunshot wound to the abdomen, loss of sensation in both legs, and hematuria (Figure 3a). From the first two findings, TraumAID derives a goal of diagnosing a compound fracture of the lumbar vertebra, and from the type of wound and the finding of hematuria it derives a goal of diagnosing renal injury. Both goals can be addressed by a single abdominal CT-scan.

Figure 3b shows the plan that is inferred from the physician's order of a lateral abdominal x-ray, which is appropriate for the goal of diagnosing a fractured vertebra, but not for diagnosing renal injury. In this situation, two points are clinically relevant: (1) the goal of diagnosing renal injury should be addressed, and (2) both goals can be addressed with the single action of an abdominal CT-scan, obviating the need to do an X-ray to address one of them. In this case a comment such as the following would be produced:

> "TraumAID suggests getting an abdominal CT-scan rather than a lateral abdominal X-ray. The former can be used both to diagnose renal injury and to diagnose a compound fracture of the lumbar vertebrae."

5 Conclusion

We have presented an approach to information delivery during *patient-centered* activities, in which a health care provider's (HCP) attention is on his or her patient and not on a computer screen. As well as potential applications to other areas of medical care, the idea of restricting the output of a decision-support system to clinically relevant information can be generalized to other *task-centered* activities, such as repair or construction, in which the person for whom the system output is intended is focused on performing a task rather than on the system.

We recognize the need to validate the approach described here, and have proposed (with Sandra Carberry, of the University of Delaware) such a study to the National Library of Medicine. In this study, we will compare two versions of our approach against graphic display of the stable portion of TraumAID's management plan on a monitor positioned in the trauma bay. In one version of our approach, critiques of physician orders (or lack thereof) will be conveyed textually to the scribe nurse, who will convey it to the other members of the trauma team at his or her discretion. In the other version, critiques will be conveyed directly to the trauma bay through synthesized speech.
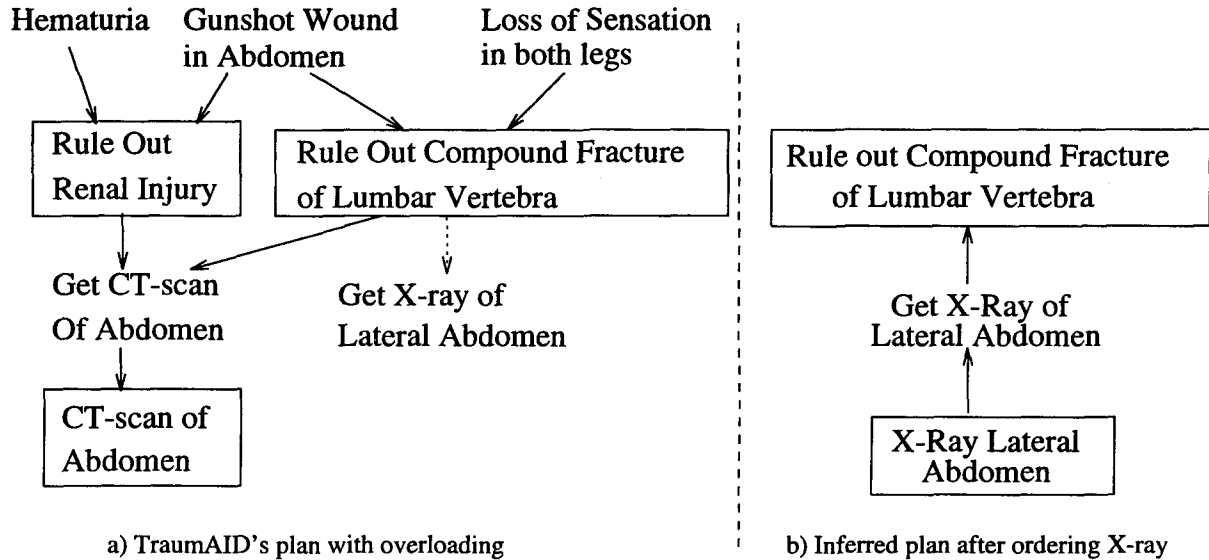
Figure 3: Inferring a plan with overloading

Even though this validation has not yet been done, we believe this paper has independent value in (1) pointing out additional constraints on language generation raised by the desire to convey information to listeners attending to something other than an computer terminal, and (2) pointing out some features of plan inference and evaluation raised by multiple goal planning in a complex domain.

References

Clarke, J.R., Rymon, R., Webber, B., Hayward, C., Santora, T., Wagner, D. and Ruffin, A. The Importance of Planning in the Provision of Medical Care. *Medical Decision Making* 13(4), October-December 1993, p. 383 (abstract).

Clarke, J.R., Webber, B., Niv, M., Rymon, R., Gertner, A. and Kaye, J. The Care of Injured Patients: An architecture of Medical Knowledge. *Der Chirurg* (Special issue on surgical decision-making), to appear April 1994.

Gertner, A. Ongoing Critiquing During Trauma Management. *Proc. AAAI Spring Symposium on Artificial Intelligence in Medicine: Interpreting Clinical Data*, Stanford CA, March 1994.

Gertner, A. Responding to Users' Informational Needs in Time-Critical Situations *Proc. 4th International Conference on User Modeling*, Hyannis, MA, August 1994.

Grice, H.P. Logic and Conversation. In H.P. Grice, *Studies in the Way of Words*. Cambridge MA: Harvard University Press, 1975. Reprinted in S. David (ed.) *Pragmatics: A reader*. Oxford: Oxford University Press, 1991, pp. 305-315,

Kautz, H. A circumscriptive theory of plan recognition. In Jerry Morgan Philip R. Cohen and Martha E. Pollack, editors, *Intentions in Communication*. Bradford Books, 1990.

Langlotz, C. P. and Shortliffe, E. H. Adapting a consultation system to critique user plans. *International Journal of Man-Machine Studies*, 19:479-496, 1983.

Miller, P. L. *Expert Critiquing Systems: Practice-Based Medical Consultation by Computer*. New York: Springer-Verlag, 1986.

Pollack, M. Overloading Intentions for Efficient Practical Reasoning. *Nôus* XXV(4), 1991, pp. 513-536.

Rymon, R. *Diagnostic Reasoning and Planning in Exploratory-Corrective Domains*. PhD Dissertation. Dept. Computer & Information Science, U. Pennsylvania, November 1993. (Appears as Technical Report MS-CIS-93-84).

Rymon, R., Webber, B. L. and Clarke, J. R., Progressive Horizon Planning – Planning Exploratory-Corrective Behavior. *IEEE Transactions on Systems, Man, and Cybernetics* 23(6). Special issue on Planning, Scheduling and Control, November 1993.

Sperber, D. and Wilson, D. *Relevance: communication and cognition*. Harvard University Press, Cambridge, Mass., 1986.

Webber, B., Rymon, R. and Clarke, J.R. Flexible Support for Trauma Management through Goal-directed Reasoning and Planning. *Artificial Intelligence in Medicine* 4(2), April 1992, pp. 145-163.