Application of Cognitive Science Methods to Psychotherapeutic Problem Solving: A Case Study and Some Theory

Rolf Pfeifer
Department of Computer Science
University of Zurich, Switzerland
and
Marianne Leuzinger-Bohleber
Department of Psychotherapy
University of Ulm, FRG

ABSTRACT

In this paper we present a case-study to demonstrate an application of concepts of knowledge representation from cognitive science and AI to problem solving in psychotherapeutic situations. In particular, a special type of frame, the so-called "Thematic Organization Point", or TOP, is used to characterize generic conflictive patterns of interaction, and to elucidate the meaning of a "psychotherapeutic interpretation". The concept of "failure-driven memory" is related to the process of evoking memories in patients. A belief systems analysis is used to explain why in some situations people are incapable of learning in spite of repetitive expectation failures. The underlying theory is summarized as a set of "Theorems". It is concluded that a cognitive science approach to therapeutic problem solving not only clarifies theoretical concepts but enables the derivation of powerful heuristics to be used by therapists in their practical work.

INTRODUCTION

Much of cognitive science research in one way or another has to deal with the notion of goals. One of the reasons is, of course, that behavior is controlled by the goals to be achieved. Another has to do with the fact that goals have a high explanatory value. Whenever we observe some kind of behavior, we typically interpret it as a plan to achieve some goal. Goals are ubiquitous in interpreting the environment, as well as in generating behavior, and therefore it seems natural to organize a knowledge base for problem solving (which is a kind of goal-directed behavior) around the notion of goals, goal-configurations and related ideas, such as plans, failures, etc. to be discussed below. These ideas will be applied to a special kind of expert behavior, namely the one of psychotherapists.

Psychotherapists are expert at interpreting unusual and unexpected behavior in their patients. For example, they search for "hidden" goals, i.e. goals the patient may not be aware of him or herself. One method frequently used to elicit a patient's knowledge structures is that of "free association" where memories related (by assumption) to the current behavior or experience are to be evoked. This method is used mainly in so-called *psychodynamic* therapies, i.e. therapeutic techniques dealing with the explanation and interpretation of behavior in terms of cognitive and emotional processes which may or may not be conscious. It is this kind of therapy on which our paper is focused. The memories thus activated can guide the way to an interpretation, i.e. they can help finding an explanation of some maladaptive or undesirable behavior. It is a basic tenet of many forms of

psychotherapy, that such an insight constitutes the basis of a learning process which can eventually lead to an elimination of certain behavioral patterns.

Psychotherapeutic theories are typically vague and at times even have a mythical touch. Cognitive science provides concepts and formalisms which should be capable of capturing at least a small portion of the phenomena involved. It is demonstrated in the sequel that not only concepts from psychodynamic theories can indeed be formalized in a transparent manner, but that this approach can directly support the practical work of the clinical expert.

After introducing the basic concepts, a case report from a psychotherapeutic treatment is presented. This case report will be analyzed in the subsequent sections with a particular focus on the therapist's hypotheses (her beliefs) about the mental representations of the patient (i.e. the patient's beliefs).

BASIC CONCEPTS

Much of the theory presented in the sequel is based on the work of the Yale Artificial Intelligence and Cognitive Science Programs (e.g. Schank, 1982). Basic familiarity with the concepts of frame theory in general, and Thematic Organization Points (TOPs) in particular, is assumed, and only those principles needed later on will be summarized as "Theorems"[1]. The theorems presented in this section deal with memory processes. They will be applied to elucidate concepts from psychodynamic therapy.

THEOREM 1: Reminding of episodes across contexts is enabled when the episodes are stored under the same TOP. The term "context" is used in its everyday sense.

Clearly, this is not the only way in which remindings can occur (they could e.g. also be visually based), but it is the most relevant one for goal-oriented and interactive behavior.

DEFINITION: A TOP (=Thematic Organization Point) is an abstract memory structure, a frame type, containing (at least) the following components: goal configuration, expectations about plans (or more generally: behavior) and outcomes, actual plans (behavior) and results, and explanation of discrepancies.

Theorem 2 deals with expectation violations.

THEOREM 2: (principle of "failure-driven memory") If an expectation generated from a frame is violated, a pointer is established from this frame to a representation of the situation in which the violation occurred. This enhances memory recall.

From Theorem 2 Theorem 3 can be derived (e.g. Schank, 1982):

[1]The term "Theorem" is not used in a rigorous mathematical sense but rather to designate theoretical concepts that have proved useful in many AI programs and for which there is considerable empirical support (e.g. Seifert et al., 1984), or heuristics on how to proceed in analyzing people´s knowledge structures.

THEOREM 3: The principle of "failure-driven memory" underlies a large class of learning processes.

The heuristic stated in Theorem 4 is a corollary of Theorem 3:

THEOREM 4: If learning processes are to be induced it is a good idea to violate existing expectations or to evoke expectations which can then be violated.

The theory presented so far can be used to elucidate the concept of a so-called "complete psychodynamic interpretation", and to explain why a psychotherapist should remain neutral towards a patient's goals and wishes if memories are to be evoked from the patient (see Boxes 1 and 2). Further "Theorems" will be introduced as needed in the case study.

BOX 1: A complete psychodynamic interpretation and TOPs

A major goal in psychodynamic therapies is to make hidden knowledge structures (e.g. goals or wishes) underlying maladaptive and undesirable behavior explicit in order to make them inspectable and potentially modifiable. Such hidden knowledge structures are frequently related to unresolved conflicts which have their origins in early childhood, and which can be observed in the patient's actual behavior. A central therapeutic mean to communicate hypotheses about maladaptive behavior to a patient is the one of a so-called interpretation which relates unresolved conflicts to behavior in different kinds of situations. According to Menninger (1958) an interpretation is only *complete* when a conflict has been understood with respect to three domains in the patient's life, namely: current therapeutic interaction, early childhood, and current life situation (work, lovers, etc.). Thus, similar patterns of behavior must be ferreted out and described in all three domains. It is exactly this similarity we are trying to capture with the (context independent) concept of a TOP. Such similarities are normally perceived by therapists in a rather intuitive and global fashion. They can be made more precise and transparent if TOPs are used for their description.

BOX 2: Therapeutic neutrality and "Failure-driven memory"

The practice of psychodynamic therapy has shown that in order to evoke memories which relate to hidden conflicts of a patient, it is appropriate for a therapist to take an attitude of neutrality: the therapist refuses to fulfill inadequate (i.e. "infantile") wishes of the patient, and thus earlier experiences come to mind and can be reflected in relation to current problem situations (see also Box 1). Wishes of a patient are considered inadequate if they imply any form of direct help or advice by the therapist. Therapeutic neutrality, as many other concepts in the psychodynamic literature, are based on intuitions and clinical experience and are not sufficiently grounded in theory. The principle of "failure-driven memory" (Theorem 2) explains why early childhood memories concerning infantile wishes can be evoked through a process of frustration in the therapeutic interaction: those memories relating to conflictive content typically have to do with expectation failures (e.g. unfulfilled wishes from a patient's distant past). Since, through expectation failures, those memories were originally tagged, their recall is enhanced if similar expectation failures occur within the therapeutic situation.

PFEIFER & LEUZINGER-BOHLEBER

CASE STUDY

Episodes from a psychotherapy of a 38 year old depressive woman, reported by Marianne Leuzinger-Bohleber: In July 1979 Gertrud, a German social worker, mother of seven children, married to a self-employed businessman, asked for an appointment at our counseling center in Zurich. She was in a deep crisis which was apparently triggered by her husband's starting his own business in the area of therapeutic instruction materials. It was her husband's wish that she give up her part-time job in a home in order to help him with his business. He was not a good businessman without her help and Gertrud was afraid that they might soon have to file bankruptcy.

In her psychotherapy she frequently had the feeling of giving herself up, that she would "never be able to free herself of me" and that she would come out of therapy as a "case". The sessions often had a tormenting quality and I couldn't reach her. Nevertheless, her marital situation had improved, among other things, through the fact that she had taken up working part time outside of her husband's firm again.

After some time she felt worse again. She complained about finding her life too taxing and that nothing was moving in her therapy. Casually she mentioned how nice it would be to have an additional cuddly silk cushion on the couch in the counseling room. I felt hard-pressed and had the impulse to go downtown and buy her such a cushion immediately.

In the following session I completely forgot about the cushion so that I couldn't even broach the subject. Gertrud did not say anything about it either but could hardly be reached in this and the following sessions. She felt distant, tacitly accusing and aggressive. But I had no idea that it could be related to the cushion.

The "cushion theme" re-emerged as Gertrud's fears of committing herself to therapy and the subject of terminating it were discussed. She explained how much she had been hurt by my forgetting the cushion and that her disappointment and her anger about the event were still present and would weigh heavily in her decision of whether to continue therapy or not. To her this had been proof that I was a cold, unempathetic person and that for me she was in fact only a "case," or a possibility to make money. This confirmed her experience once more that she had no bearing in the world. She felt exploited and didn't trust me any more.

This problem could be worked on therapeutically in the following weeks as she was beginning to feel that crucial infantile memories were being activated, though they were not conscious at the time. This lead to her decision to continue therapy in order to "go after" the core conflict.

After working on the cushion experience a host of memories emerge. We only render one example which will be used in our subsequent analysis: When she was about four or five (in 1946) Gertrud had to accompany her mother when at night she returned to the occupied zone to retrieve certain objects they had left behind, e.g. a typewriter. Her mother used her as a sort of "protection" against the Russian occupational forces. During these "excursions" she was scared to death but was not even allowed to cry.

Summary of report: In this paper the following key scenes are discussed:
the therapeutic situation: the patient communicates that it would be nice to have a silk cushion and doesn't get it ("cushion scene"),

- the patient's current real-life situation: the patient should help her husband with his business ("husband's business scene"),
  an early childhood experience: the patient had to accompany her mother into the occupied zone to "protect" her against the occupational forces ("occupational forces scene").

ANALYSIS OF KEY SCENES

The goal of this analysis is to make explicit the patient´s beliefs and memory representations, in particular the ones controlling her behavior. We begin with the "cushion scene". From the patient's utterance we infer that the patient has the goal to have such a cushion. This is not by logical necessity, and the attribution of goals to an actor is indeed a possible source of error. Moreover, we interpret her utterance as part of a plan to achieve her goal: the therapist is to be her "agent" (Schank & Abelson, 1977). The patient expects the therapist to fulfill her wish. However, this expectation is violated. As we find out later, the patient had an explanation of why her wish was not fulfilled, namely that the therapist is cold and non-empathetic. This preliminary analysis is summarized in Figure 1.

Patient´s structure

GOAL: SELF has silk cushion.

PLAN: SELF uses T as an "agent". SELF communicates goal to T.

EXPECTED RESULT: SELF has silk cushion.

ACTUAL RESULT: SELF does not have silk cushion.

EXPLANATION OF DISCREPANCY BETWEEN EXPECTED AND ACTUAL RESULT: T does not respond to the needs of SELF since T is cold and unempathetic.

FIGURE 1: Patient´s representation of the "cushion scene": preliminary hypothesis. "SELF" refers to the patient, "T" to the therapist.

For the purposes of the present discussion we will ignore the therapist's own goals and plans for the therapy and concentrate on the ideas she has about the patient. From the utterance of the patient she infers that the patient has the goal to have a cushion. But the patient does not explicitly talk about this wish and report her related thoughts, as could be expected in therapeutic settings. This implies that the therapist's expectations are violated: the patient wanted her to actually buy her a cushion rather than work out an appropriate interpretation. This discrepancy is also reflected in the therapist's emotional reaction. A major goal for the therapist now is to find appropriate explanations.

Frequently, when our expectations are violated and there is no adequate explanation of a person's behavior available this may be due to the fact that our hypotheses about that person's goals are inappropriate. In our example, the patient may never have had the goal to have an additional cushion. Theorem 5, adapted from Wilensky (1983), gives us a clue as to what some alternative goals might be.

THEOREM 5: The execution of a plan frequently activates "dormant" goals, i.e. goals that would otherwise not have been activated. Of particular interest are the so-called "preservation goals".

Patient's structure

T'S GOAL: To be a successful, financially independent therapist.

T'S PLAN: Use SELF as an "agent".

leads to

PRESERVATION GOAL: SELF does not want to be merely a "case" (i.e. an "agent") for the benefit of T.

PLAN FOR PRESERVATION GOAL: SELF wants from T some kind of behavior which does not conform to the rules of therapeutic technique (provide SELF with silk cushion). Mention goal in conversation.

EXPECTED RESULT FOR SELF: T gets her the cushion.

ACTUAL RESULT FOR SELF: T does not get SELF the cushion. Strong negative emotions. SELF is only a "case".

EXPLANATION OF DISCREPANCY BETWEEN EXPECTED AND ACTUAL RESULT: T deliberately ignores preservation goal of SELF. Thus SELF is merely an "agent" of T.

ACTUAL RESULT FOR T: T achieves goal: T makes money and is successful.

FIGURE 2: Patient's representation of the "cushion scene": more elaborate hypothesis. "SELF" refers to the patient, "T" to the therapist.

If we look at the intensity of the reactions of both the patient and the therapist to the "cushion scene" (the patient hardly talks to the therapist for a long time; the analyst "forgets" about it) we infer that there must be something major at issue. The major activity is the patient's involvement in the therapy. Her plan is to use the therapist as an agent to help her solving her problems. If we look through the case report for possible preservation goals which are activated by her involvement, we find a statement to the effect that she does not want to be simply a "case". Applying this idea to the "cushion scene" we find a different interpretation of the patient's utterance that it would be nice to have a cuddly silk cushion on the couch. Had the therapist assented to the implied wish of the patient, this would have been an action on the part of the therapist which lies outside of the range of therapeutic (neutral) behavior. In that case -- in the patient's beliefs -- the patient would not have been treated merely as a "case" but as a "real person"; the therapist would have fulfilled her wish even though this did not conform to the rules of psychotherapeutic technique. The utterance concerning the cushion can now be interpreted as a (clever) plan to achieve the preservation goal of remaining an independent "real" individual and not merely being a "case". Of course, such a plan could only be conceived by someone with profound knowledge of psychotherapeutic technique, a precondition the patient clearly fulfilled. It is interesting to note that the patient, in order to get her preservation goals met, tries to involve the same person as an agent who the patient holds responsible for her preservation goal in the first place (since the therapist is the one who uses her as an agent). This more elaborate analysis is summarized in Figure 2. Again, it is a hypothesis of the therapist about the

patient's mental representation. This structure is non-trivial and goes beyond a simple goal-subgoal analysis. As a matter of fact, we are dealing with a rather tricky and uncomfortable pattern of interaction.

If this interpretation of the patient's behavior is appropriate and if it is communicated to the patient, then corresponding memories should be activated from the patient's distant past (early childhood) and from her recent past (her real-life situation). And indeed, when this information could be made explicit the patient started reporting her early childhood memories. Since in early childhood, for example, there were no therapists, and no silk cushions, an abstraction has to be made from the current context, with the constraint that the relevant information concerning behavior and interaction is preserved. As mentioned earlier, these commonalities are captured by TOPs. The TOP corresponding to Figure 2 is shown in Figure 3. We will call it "GA-UP" (Goal Pursuit with Agency - Unsatisfiable Preservation Goal).

Patient's TOP

O'S GOAL: Important goal.

O'S PLAN: Use SELF as an "agent".

leads to

PRESERVATION GOAL: SELF does not want to be an "agent" of O. O is an "agent" for SELF.

EXPECTED RESULT FOR SELF: O understands and achieves goal of SELF.

ACTUAL RESULT FOR SELF: SELF is an "agent". Strong negative emotions.

EXPLANATION OF DISCREPANCY BETWEEN EXPECTED AND ACTUAL RESULT: O's goal has such high priority that O deliberately ignores preservation goal of SELF.

ACTUAL RESULT FOR O: O achieves goal.

FIGURE 3: Patient's TOP, called "GA-UP" (Goal Pursuit with Agency - Unsatisfiable Preservation Goal). "SELF" refers to the patient, "O" to the person the patient is interacting with (the "other").

For a complete psychodynamic interpretation (see Box 1) we need also a representation of an early childhood experience and of an episode concerning the patient's current life. With the TOP "GA-UP" it is straightforward to find the representations of the "husband's business scene" and the "occupational forces scene". The former is shown in Figure 4, the latter is left as an exercise to the reader (or see Pfeifer & Leuzinger-Bohleber, 1986).

Patient's structure

H'S GOAL: To be an independent, financially sound businessman.

H'S PLAN: SELF must help out in H's store.

leads to

PRESERVATION GOAL: Maintain independent life style (keep job in home). SELF does not want to be merely an "agent" to run H's business.

PLAN FOR PRESERVATION GOAL: SELF wants H to support preservation goal. SELF communicates goal to H.

EXPECTED RESULT FOR SELF: H helps SELF keep independent life style.

ACTUAL RESULT FOR SELF: H interprets needs of SELF as exaggerated. H does not support wish for independence of SELF. Strong negative emotions.

EXPLANATION OF DISCREPANCY BETWEEN EXPECTED AND ACTUAL RESULT: H does not understand the importance of independence for SELF. H exploits SELF to achieve his own goals.

ACTUAL RESULT FOR O: H achieves goal: Business of H improves.

FIGURE 4: Patient's representation of the "husband's business scene". "SELF" refers to the patient, "H" to her husband.

REPETITIVE BEHAVIOR

Frames can be used to interpret the environment as well as to generate behavior. From the patient's reports it is obvious that the TOP "GA-UP" is often activated and controlling her behavior, and each time it produces undesirable negative results. One question that immediately comes to mind is why the patient is apparently incapable of learning from her expectation failures, although they occur repetitively. This persistent repetition of some undesirable behavior in different contexts is called "repetition compulsion". Theorem 6 states one frequent reason.

THEOREM 6: If an individual is incapable of learning from her expectation failures, this may be due to the fact that a structure (a frame) different from the believed one is controlling her behavior.

Theorem 6 can be directly applied to our case study. The reason why the patient is not capable of changing her behavior, although she perceives it as undesirable, is that in her (conscious) belief, she is applying a simple planning frame (called, say, "use agent") for engaging someone as an agent to achieve her goals (and this planning frame normally works - in situations not described in the case report), whereas in "reality" she is activating the more intricate frame "GA-UP". "GA-UP" is activated whenever the following triggering conditions are fulfilled: the situation contains a person who is important to the patient; this person is pursuing an important goal (in the patient's beliefs, not in "reality"); and the patient is used as an "agent" in the plan of that person. These conditions are

fulfilled for all three key scenes described. But the patient cannot discriminate between situations in which she is applying the plan "use agent" and situations in which she is using "GA-UP", for the simple reason that she has no awareness of "GA-UP". Since her explanations of expectation failures are based on the wrong assumptions, she cannot change her behavior appropriately. One fact that adds to the severity of the problem is that "GA-UP", like "use agent" is context independent and can be instantiated in many situations.

There are two main reasons for believing that our analysis is appropriate. First, the patient's emotional reactions to her expectation violations are much too strong if the plan had simply been "use agent". The intensity can only be explained in relation to the negative and strong early childhood experiences. This relation is captured by "GA-UP" which relates the memory representations of early childhood to current experiences. Second, we had the opportunity to follow her therapeutic development. The patient experienced the recognition of "GA-UP" as a profound insight, and from that point in time, her behavior started to change: the preconditions for a learning process were met (for more detail, see Pfeifer & Leuzinger-Bohleber, 1986).

DISCUSSION

Knowledge representation: The combination of the idea of belief systems and the Yale concepts for knowledge representation enabled us to explain why certain types of behaviors can be transferred to a variety of different domains, and why some of them are so persistent even if they are perceived as undesirable. Moreover, it was possible to devise heuristics which may be used in therapeutic problem solving, and to make some of the rather opaque concepts from psychodynamic therapy more transparent. The powerful concept of a TOP was applied to the representation of patterns of interaction in different domains (e.g. Box 1). It is suggested that it may be a good heuristic for a psychotherapist to actively try to work out TOPs since they not only characterize conflictive interactions, but they may also help to activate pertinent memories (so-called "strategic reminding"; Schank, 1982).

Evocation of memories: Using the concept of "failure-driven memory" it was demonstrated why the therapist should remain neutral towards a patient's goals and wishes: memories can be evoked in this way. And memories are precious sources of information for therapeutic problem solving.

Related work: There have been many frame approaches to knowledge representation in AI and cognitive science research but for our own purposes we found the emphasis on expectations and explanations of expectation failures, as well as their relationship to memory processes (Schank, 1982) most convincing. Similar concepts were used, for example, by Dyer (1983), Kolodner et al., (1985), and Lehnert (1981) mainly for the representation of complex patterns of interaction. A closely related study is the one of Teller & Dahl (1981). They also use a frame concept to describe situations in psychotherapy transcripts. In their paper, similarity between situations is defined in terms of the rather technical notion of "frame overlap". In contrast, the concept of a TOP as a specific frame type provides similarity relations on the basis of a general theory of cognitive processes.

REFERENCES

Dyer, M.G.: *In-depth understanding*. Cambridge, Mass.: MIT Press, 1983 (originally published, 1982)

Kolodner, J.L., Simpson, R.L. Jr., Sycara-Cyranski, K.: A process model of case-based reasoning in problem solving. In: *Proceedings of the 9th International Joint Conference on Artificial Intelligence,* 1985, 284-290.

Lehnert, W.G.: Plot units and narrative summarization. *Cognitive Science*, 1981, *5*, 293-331.

Menninger, K.: *Theory of psychoanalytic technique*. New York: Basic Books, 1958.

Pfeifer, R., & Leuzinger-Bohleber, M.: Applications of cognitive science methods to psychoanalysis: A case study and some theory. *International Review of Psycho-Analysis,* 1986, *13,* 221-240.

Schank, R.C.: *Dynamic memory*. Cambridge, UK: Cambridge University Press, 1982.

Schank, R.C., & Abelson, R.P.: *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

Seifert, C.M., McKoon, G., Abelson, R.P., & Ratcliff, R.: Memory connections between thematically similar episodes. Yale University Cognitive Science Technical Report #27, 1984.

Teller, V., & Dahl, H.: The framework for a model of psychoanalytic inference. In: *Proceedings of the 7th International Joint Conference on Artificial intelligence,* 1981, 394-400.

Wilensky, R.: *Planning and understanding*. Reading, Mass.: Addison-Wesley, 1983.