Where Do Goals Come From?

Jaime G. Carbonell
Carnegie-Mellon University
Pittsburgh, PA 15213

Abstract

Theories of rational behavior embodied in cognitive models of problem solving, planning, and plan interpretation typically presuppose that the planning agent is given *a priori* one or more goals to pursue. Thereupon, rational behavior consists of planning and carrying out a sequence of actions in order to achieve the most important active goals. This paper argues that a complete cognitive model must necessarily incorporate *the process of acquiring goals* whether in reaction to perceptions of external events, in response to internal physiological or psychological states, or by other less direct means. An initial categorization is made of various mechanisms that can give rise to goals in an individual planner.

1. Introduction

The AI literature abounds with models of problem solving, planning and plan interpretation (e.g., GPS [11], STRIPS [6], NOAH [14], PAM [18], BELIEVER [16], TALESPIN [10], POLITICS [5, 3]). Although these models differ in terms of the specific cognitive phenomena simulated, in terms of their internal structure, in terms of their representation formalisms, and in terms of their theoretical motivations, it is striking that they all share one central hypothesis: Each and every system is heavily dependent upon the presence of one or more goals attributed to the active problem solving agents or planners. In essence, each planner or problem solver incorporates an implicit theory of rational behavior based upon the assumption that all actions are preformed in service of explicit, realizable goals. Therefore, rational behavior for a planning system consists of formulating a sequence of planned action to achieve a set of goals. In the case of story interpretation, the assumption of rationality applies to the characters, and the task of the understander becomes one of divining their goals by reconstructing corresponding plans from sequences of observed events.

Hence, under these models of planning and plan interpretation, rationality becomes synonymous with intelligence. Or, as Newell defines it: Intelligence is the ability to bring knowledge to bear in the pursuit of goals [12]. Wilensky [19] also articulates the notion that all intelligent action ensues from the pursuit of multiple goals, including the resolution of internal goal conflicts by the spontaneous creation and subsequent pursuit of *metagoals*. The implicit centrality of goals becomes more evident when one considers some attempts at modeling affect or idiosyncratic behavior. For instance, Lehnert's affect states [9] in story interpretation, and recent work on modeling emotions [1, 13] rely on mechanisms to detect goal frustration or goal achievement. My earlier work on modeling ideological belief and certain aspects of human personality traits relies even more heavily on the presence, pursuit and attribution of different types of goals to planning agents [2, 5].[1]

2. Goal Generators in Integrated Cognitive Models

If goals are central to all effective AI theories of intelligence, the natural question arises: *Where do goals come from?* Whereas taxonomies of goals [15], relations among the goals of an individual [5, 19], and methods of planning to achieve goals are all significant aspects of the study of goals, the key notion of *what cognitive, physiological or social mechanisms give rise to goals* has been largely glossed over by AI researchers. An AI program, whether planner or problem solver, does nothing until an external entity (such as the programmer) provides it with a goal to pursue, whereupon the program single-mindedly strives to find an effective plan for that goal, and regardless of success or failure, resumes idling indefinitely after the solution attempt. Clearly, any complete cognitive model must generate its own goals. Philosophical debate on issues of free-will vs determinism notwithstanding, all intelligent beings exhibit some measure of internal motivation and ability to respond to unexpected situations in the external environment.

The type of integrated cognitive model I envision would contain a goal generator that would monitor continuously the external environment and its internal state as a background process, and hence it would *notice* if it is getting hungry or tired, or that an external threat is imminent, bringing these issues (perhaps as interrupts) to the attention of the conscious "rational" processor, which then may decide to generate new goals, reprioritize existing goals, or ignore the interrupts. Essentially, the continuous monitoring of possible sources of goals necessarily forces one to face the issue of focus of attention, an issue that can be safely ignored only as long as an external entity provides all goals and thereby limits distracting factors. In fact, the single-minded pursuit of a small set of externally imposed goals determined *a priori* obviates the need to refocus attention dynamically as no unforeseen happenings will be noticed. Consider a present-day AI planning system deciding, for example, how to stack blocks. When faced with an external threat or a greater need, it will not have the sense to abandon or postpone its present task, generate and pursue a more appropriate goal, and thereby change the current focus of attention.

Rather than attempting the formidable task of characterizing the space of plausible cognitive models capable of directing their own attention, and responding to changing events by generating their own set of appropriate goals, let us focus on the more tractable subproblem of exploring various mechanisms capable of generating goals dynamically.[2] From a psychological standpoint, an obvious source of goals is the internal physiological state of the planning agent: Hunger leads to the goal of satiation of hunger; physical exhaustion leads to a desire for rest. From an AI standpoint, an equally obvious source of goals is the planning system itself generating subproblems, with the associated goal of solving the subproblem. For instance, an AI planner may decide that, given the externally imposed goal of "satiate hunger", it should first locate food, then transport itself to that location, then ingest the food. Each of these steps, if not immediately executable in the external world, generates a subgoal requiring additional planning (e.g., locating food generates the subgoal of knowing the location of the food, which then may lead to searching or asking, etc.) There are, however, more complex sources of goals. Schank and Abelson postulate a set of *themes* as goal generators whose internal structure remains a virtual black box. For instance, the *love theme* generates the goal of protecting one's loved ones. Unlike other aspects of Schank and Abelson's theory of representation and understanding, their treatment of themes does not provide a very satisfying analysis, in that it neither postulates a computational mechanism for how these themes operate or are

---

[1] In this argument I do not mean to imply that all theories of emotion or even theories of human intelligence interesting to AI practitioners are necessarily based on goals and their unrelenting pursuit. I am merely noting that theories precise enough to result in operational process models (e.g. AI programs) incorporating significant aspects of human cognition have *thus far* been dependent on goals and the implicit principle of rational behavior.

acquired, nor does it attempt exhaustive coverage or broad sampling of cognitively plausible goal generators. Here, we pursue the latter goal with the longer range objective of eventually developing computational mechanisms that give rise to goals in the context of a complete cognitive model.

3. Towards a Taxonomy of Goal Generators

Let us again pose the central question: *Where do goals come from?* However, rather than examining the literature for possible answers as I attempted above, let us enumerate and categorize possible goal generators in humans. It appears that the following general categories cover a large range, if not the entire space of goal generators:

1. Internal physiological state changes
2. Mental (e.g., emotional or attitudinal) state changes, possibly accompanied by, or resulting from physiological state changes
3. Knowledge state changes
4. Perceptions of changes in the external world
5. Socially imposed goals or constraints on the individual
6. Instrumentality (i.e., goals generated purely in service of other goals)

Examining this list, several observations become readily apparent:

- General coverage is indeed attained, in the sense that goals typically attributed to people can be coerced into a combination of one or more of the categories above.
- This list is of very little use in developing a process model, as it lacks commitment to any *fine-structure detail*.[3] Generality is not the only metric one should apply in judging the utility of a theoretical concept.
- The classification itself does not necessarily suggest that a uniform mechanism operates within each category giving rise to the set of goals thus grouped together. Therefore, if the analysis is to be useful in constructing a predictive, psychologically plausible, process model, the categorization must be motivated more strongly by the *processes* that operate in generating the classes of goals grouped together.

Bearing these concerns in mind, let us construct a more detailed categorization motivated by commitment to finer-structure detail of the processes that generate goals, and let us place less emphasis on global generality at this stage of the investigation. In the taxonomy of goal generators presented below, the hierarchical structure is meaningful, as are the suggested mechanisms, but the order in which the categories are listed is quite arbitrary.

1. **INSTRUMENTALITY**

   a. **Direct instrumentality** -- Given a higher level goal, subgoals are generated by the planning or problem solving process whenever a step in the plan to achieve the higher level goal is not directly realizable, and hence requires additional directed planning. These goals correspond to Schank and Abelson's "delta goals" [15].

   b. **Derived or indirect instrumentality** -- Secondary goals instrumental to the achievement primary goals arise through several mechanisms in addition of strict subgoal instrumentality, to wit:

      i. In the process of planning to achieve more than one primary goal, conflicts may arise among active goals of the planner giving rise to *metagoals* [19] of resolving the internal goal conflict in order for the planner to achieve all (or the most crucial subset) of his primary goals. Typically these conflicts are based on resource limitations, including limitations on the time that the active planner can devote to a particular set of tasks.

      ii. In the *counterplanning process* [4, 5], instrumental goals of assuring that an adversary cannot (or will not) thwart an otherwise viable plan arise frequently. These are not true subgoals, in that they may play no role in achieving the primary goal, but rather may be directed at misleading, diverting or negotiating with potential adversaries.

      iii. *Goal subsumption states* [19] arise when a primary goal recurs frequently, or many primary goals share a common instrumental subgoal. In essence, a subsumption state facilitates the achievement of many instances of primary or instrumental goals. Hence, the achievement of a desired subsumption state becomes a goal in itself. An instance of a subsumption state is having a steady income, thus facilitating any goals requiring money, and aiding social-status goals as well. Similarly, establishing an alliance to aid in future mutual fulfillment of different primary goals, or terminating an adversary relation can be considered subsumption goals [5].

      iv. *Optimization* of a plan, or *saving mental effort* while planning could be construed as indirect instrumental goals to the primary objective.

2. **INTERNAL DRIVES** -- these may be considered psychologically inate goals in an individual

   a. **Cyclic physiological drives** -- these are goals generated in response to internal physiological states that change with a certain periodicity. A cognitive model may treat the mechanism that generates basic drives of this sort as a black box. Schank and Abelson label these "Sigma goals". A partial enumeration of cyclic physiological drives includes:

      i. Satiation of hunger

      ii. Satiation of thirst

      iii. Desire for rest or sleep

      iv. Desire for sexual activity

   b. **Non-cyclic physiological drives** -- these occur primarily in response to adverse changes in the environment, and perhaps should also be considered as black boxes when constructing a cognitive model. These goals have no correlate in the Schank and Abelson taxonomy. A representative sampling includes:

      i. Self-preservation (in response to overt threats)

      ii. Protection of one's offspring (again in response to overt threats)

      iii. Seeking warmth (if the external temperature drops)

      iv. Satisfying curiosity (e.g., in response to unexpected external events)

      v. Seeking companionship (in its absence)

3. **SOCIAL GOALS** -- These are goals that arise by virtue of interaction with other members of the species.

   a. **Semi-autonomous social dynamics** -- these goals

---

[2] The reader is referred to the "World Modeller's Project" [8, 7] for a discussion of a general experimental system that simulates a reactive environment in which one may build simple planning systems that must cope with changes in the environment. Such a system is an experimental tool that expedites research and sheds light on significant problems not heretofore investigated in the appropriate context. (Such problems include the topic of this paper.)

[3] Sloman argues convincingly that evaluating a theory based solely on breadth of coverage and predictive generality ignores issues of internal structure and commitment to detail, which often differentiate useful theories from general truisms [17].

appear to require no explicit learning, but arise only if an individual interacts with other members of the species. Again, these goals have no direct correlate in Schank and Abelson's taxonomy. Types of semi-autonomous social goals include:

i. Simple social ambition (e.g., become the king of the hill, or the leader of the pack, or the respected medicine man)

ii. Property ownership, acquisition and protection from others (There can be no meaning to ownership without the notion of restricting access to others of the objects owned.)

iii. Protection of others within the social group from external threats (This clearly goes beyond protection of self or biological offspring)

iv. Protection of the nature and makeup of the social group itself (e.g. from other members of the species who may pose no threat to individuals within the social group, but pose a threat to the established social order)

v. Jealousy, wanting something merely because another member of the social group has acquired it

vi. Avoid banishment by the social group

b. **Socially taught or imposed goals** -- unlike the previous category, these goals vary across social groups within the species, and therefore must be learned by individuals (from observation of more mature members of the social group, or by direct instruction). Here I defer to anthropologists or social psychologists to provide a more comprehensive list; the following is meant as an illustrative sample:

i. Abide by the formal and unwritten laws of the society

ii. Live according to the ethics and morals adopted or imposed by the society on the individual

iii. Contribute to the communal wealth and well being (in some societies)

iv. Seek to attain those qualities that comprise a metric of status in the society (wealth, power, respect, wisdom, notoriety, etc. depending on the particular society)

4. ENJOYMENT GOALS -- these correspond roughly with Schank and Abelson's "E-goals".

a. **Direct (physiological) pleasurable experience** -- these goals overlap substantially with cyclic and other physiological goals discussed earlier; the central distinction is based on the circumstances in which they arise (e.g., the motivation to walk into a hot tub or a steam bath differs from the motivation to seek shelter in frigid weather, although the resulting goal states overlap in terms of the physical state change sought).

i. Physical exertion for pleasure (as opposed to exertion instrumental to other primary goals), such as exercise, some forms of children's play, etc.

ii. Direct sensual gratification (such as eating for pleasure in "gourmet" dining, tactile gratification, etc.)

iii. Aesthetic gratification (such as enjoying a painting, a sunset, a concert, a good novel, etc.)

b. **Derived psychological pleasure** -- satisfaction of most non-trivial goals yields a measure of resultant pleasure, but some goals appear to be caused by no internal or external reason other than experiencing this measure of indirect pleasure. For instance:

i. Vicarious pleasure (role playing, identification with characters in movies, novels or sporting events, etc.)

ii. Acquisition of knowledge for its own sake, when the knowledge is not instrumental to any primary goals, nor is its presence a realistic subsumption state (e.g., assorted trivia, half of the features stories in newspapers and magazines that bear no impact on any conceivable goal of the reader, intellectual curiosity, etc.)

iii. Acquisition of objects for their own sake (For instance, most stamp and coin collectors are not primarily motivated by the prospect of making money from their collections, but rather amassing and classifying their precious objects becomes an end in itself.)

5. MENTALLY-DERIVED GOALS -- these are goals resulting from deliberate reasoning processes, including:

a. Goals arising from mentally deduced information (as opposed to directly observed information). These goals may bear similarity in content with previous goals, but not in their method of inception (such as deciding that the disturbance in the campsite could have been caused by a grizzly bear, and hence activating the self-preservation goal).

b. Goals arising from the result of purposeful reasoning (such as deciding on a particular career to pursue after much thought). These are not instrumental goals, but often long-range personal-objective goals.

4. Concluding Remark

The goal categorization above, however imperfect or incomplete, is offered as an initial step towards developing effective models of the goal acquisition process, and thereby eventually creating more complete models of human cognition. Subsequent to the postulation of a particular taxonomy motivated by plausible sources of the various classes of goals, I intend to focus on modeling explicitly a planning agent that acquires its own goals and refocuses its attention in an interrupt-driven manner. The World Modellers project offers an amenable environment in which to create progressively more complex, cognitively plausible models that interact with a simulated environment.

5. References

1. Bower, G. H. & Cohen, P. R., "Emotional influences in memory and thinking: Data and theory," in *Affect and cognition: The 17th Annual Carnegie Symposium on Cognition*, M.S. Clark & S.T. Fiske, ed., Erlbaum, Hillsdale, N.J., 1982.

2. Carbonell, J. G., "Towards a Process Model of Human Personality Traits," *Artificial Intelligence*, Vol. 15, No. 1,2, november 1980 , pp. 49-74.

3. Carbonell, J. G., "POLITICS: An Experiment in Subjective Understanding and Integrated Reasoning," in *Inside Computer Understanding: Five Programs Plus Miniatures*, R. C. Schank and C. K. Riesbeck, eds., New Jersey: Erlbaum, 1981.

4. Carbonell, J. G., "Counterplanning: A Strategy-Based Model of Adversary Planning in Real-World Situations," *Artificial Intelligence*, Vol. 16, 1981 , pp. 295-329.

5. Carbonell, J. G., *Subjective Understanding: Computer Models of Belief Systems*, Ann Arbor, MI: UMI research press, 1981.

6. Fikes, R. E. and Nilsson, N. J., "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving," *Artificial Intelligence*, Vol. 2, 1971 , pp. 189-208.

7. Hood, G. and Carbonell, J. G., "The World Modelers Project: Constructing a Simulated Environment to Aid AI Research," *Proceedings of the Thirteenth Annual*

Pittsburgh Conference on Modeling and Simulation, 1982, Pittsburgh, PA.

8. Langley, P., Nicholas, D., Klahr, D. and Hood, G., "A Simulated World for Modelling Learning and Development," *Proceedings of the Third Annual Conference of the Cognitive Science Society*, 1981.

9. Lehnert, W.G., "Affect units and narrative summarization," Tech. report 179, Yale Univ., Dept. of Computer Science, May 1980.

10. Meehan, J. R., *The Metanovel: Writing Stories by Computer*, PhD dissertation, Yale University, Sept. 1976.

11. Newell, A. and Simon, H. A., *Human Problem Solving*, New Jersey: Prentice-Hall, 1972.

12. Newell, A., "The Knowledge Level," Tech. report, Dept. of Computer Science, Carnegie-Mellon University, 1981, CMU-CS-81-131.

13. Pfeifer, R., "Cognition and emotion: an information processing approach," Tech. report 436, Dept. of Psychology, Carnegie-Mellon University, May 1982, C.I.P. Working Paper.

14. Sacerdoti, E. D., *A Structure for Plans and Behavior*, Amsterdam: North-Holland, 1977.

15. Schank, R. C. and Abelson, R. P., *Scripts, Goals, Plans and Understanding*, Hillside, NJ: Lawrence Erlbaum, 1977.

16. Schmidt, C., Sridharan, N. and Goodson, J., "The Plan Recognition Problem," *Artificial Intelligence*, Vol. 11, No. 2, 1978, pp. 45-83.

17. Sloman, A., *The Computer Revolution in Philosophy: Philosophy, Science And Models of the Mind*, Harvester Press, 1978.

18. Wilensky, R., *Understanding Goal-Based Stories*, PhD dissertation, Yale University, Sept. 1978.

19. Wilensky, R., *Planning and Understanding*, Addison Wesley, Reading, MA, 1983.