CS 786G *Cognitive Modeling*
Week 3 Discussion Paper
Modelling Intention in SOAR

Steve Woods
89800462

May 26, 1995

Newell purports that SOAR is both a candidate unified theory of cognition, and an implementation or exemplar of that theory. The justification of this claim is the belief that the actual specification of any unified theory of cognition is also a specification of an architecture which is essentially the heart of the theory. SOAR is something of a preliminary candidate, demonstrating neither emotion nor dreams, features commonly associated with human cognition. While we may not currently associate every possible aspect of human "mind" with an analogy in SOAR, there are certainly many features *necessary* to any cognitive entity that may be examined, at least analagously, in SOAR. In this short essay we will examine one such characteristic frequently considered by AI researchers and philosophers as *necessary* to cognition, *intention*.

Essentially, Newell sees the psychology of SOAR as existing in the way in which SOAR utilizes encoded knowledge about the external world while selecting actions in the service of goals. In order to discuss *intention*, we must first define what we mean, and if we wish to compare with human "mind" we must first have a characteristic considered to be of "mind" to examine. We define *intention* as the adoption of a high level goal, and the subsequent "focus" of effort on this goal, until either the goal is satisfied, the goal has been (temporarily or permanently) identified as unlikely to produce the desired result after a certain amount of effort, or the situation has changed to the extent that intention changes to accomodate the new factors. Intention, therefore, can be seen as both deliberative and introspective in the sense that a continuous process of evaluation of current activity is undertaken, with either a positive (continue) or negative (form a new intention) decision being made. Intention regulates behaviour that seeks to link known causes to desired effects. If we allow that any cognitive entity must possess and demonstrate intentionality in order for us to recognize cognition, then SOAR must possess such an analogy if it is to be a satisfactory model and architecture of cognition.

SOAR is a general problem solving system. Specifically, SOAR is a solver that "senses" an external environment defined in terms of a given set of perceptual systems, forms a model of this environment internally with respect to *encoding* production rules, derives appropriate intended action or actions according to *cognitive production rules* using the internal environment encoding, and then operates a given set of world-effectors according to existing *decoding* production rules. But what problem is SOAR solving ? What gives SOAR the *intention* to solve anything at all ? SOAR's intention to solve a particular problem derives from a *elaboration phase* that attributes relative levels of *preference* to goals, problem spaces, states and operators operates on a *context stack* according to the firing of particular production rules. A *decision cycle* arbitrates amongst these preferences and chooses one goal, problem space, state and operator combination to effect.

It is easy to envision domains in which high level goals may be solved through several incompatible approaches. If we view each of these approaches as a specific instance of the general "goal solution method", then we may imagine the existence of several sets of production rules, one specific to each solution approach, and complete with preferences that are relevant only within the context of the specific approach for which it was intended. SOAR is free to choose any of the methods according to production preferences at the "goal solution method" level of the context hierarchy, and then invoke the more specific productions lower in the hierarchy. Further, SOAR may at some point "prefer" to move back up to the "goal solution method" level as a result of some accumulated difficulties (derived by, or recognized by other production rules) at the specific level, and subsequently choose another solution method. Consequently, in cases where very specific intermediate preferences are not well specified, SOAR's behaviour could easily degenerate into a kind of thrashing where effort is spent on one specific subgoal for a time, and then on another for a time, thus far delaying arrival at any solution at the higher level indefinitely. In such a case it would appear that the system was lacking deliberativeness, and consequently lacking intelligence.

However, Newell observes that the ability to be "distracted" is also a recognizable feature of human cognition in some instances, and that SOAR demonstrates this through it's ability at any point to change its "mind" by changing any part of its goal, problem space, state, operator tuple and essentially wander away from the intent of its previous task, or even of its "higher level" task, since the context stack is actually a hierarchy wherein each level of the stack represents subgoals of the level above. But where is the "stick-to-it-ive-ness", the directional intertia pushing SOAR forward to solve one goal before it wanders away to some other, potentially distantly related goal ? The only knowledge about "how" to act for SOAR lies encoded in its production rules, which combine to create a kind of emergent behaviour, in terms of all three of "sensing", "cogitating" and "effecting".

SOAR's intertia lies in *intermediate preference rules*[1] Competing approaches to solving an identical subgoal may have no clear preference between them *per say*, however, the existence

[1]While these preference rules are not specifically discussed in Chapter 4 of Newell's paper, their existence is implied in SOAR's traversal from abstract problem spaces and goals to specific spaces and subgoals.

of a production rule that simply bounded the effort spent on one problem space of this type by decreasing preference for this space in favor of competing spaces at an arbitrary rate would essentially endow SOAR with a measure of "stick-to-it-ive-ness" determined by the rate of declining preference. Now, SOAR's psychology can differ significantly and to obviously different degrees of effectiveness according to this rate for a particular problem type.

SOAR is consequently capable of regulating behaviour when selecting or linking causes (or productions) to desired effects (goals), according to Newell's interpretation of psychology. It can deliberate, by executing a decision process amongst competing goals, spaces, states and operators, and further, is capable of focusing on one or more of each at any level of the context space according to other general or specific rules for allocating preference. While this behaviour would appear to be analagous to *intention* according to our narrow and arbitarary definition, we are left with a somewhat ambiguous result if our definition is lacking. Is SOAR's intentionality really *how* a human being embodies or defines intention ? Almost certainly not. Can SOAR demonstrate instances of behaviour using production rules and preferences so as to guide and focus itself on problem aspects much as "intention" guides humans in their cogitation ? Yes, for limited, perhaps artifical cases. Is there really a difference that matters ?

References

Boden, M. (1989), *Artificial Intelligence in Psychology : Interdisciplinary Essays*, The MIT Press.

Newell, A. (1990), *Unified theories of cognition*, Harvard University Press, Cambridge, MA, chapter 4, pp. 158–200,226–234. ISBN 0-674-92099-6.

Thagard, P. (1994), Introduction to cognitive science, Manuscript in preparation.