THE AUTONOMOUS, RATIONAL DESIGN AGENT

GREGORY J SMITH AND JOHN S GERO
Key Centre of Design Computing and Cognition
University of Sydney
{g_smith, john}@arch.usyd.edu.au

1. Introduction

Designing begins with abstract concepts and goals and proceeds towards the concrete. Conceptual design, especially, is characterised by abstractness and an incomplete understanding of the problem and/or solution (Gero 1998). The abstract concepts in which conceptual design goals are expressed are themselves situated abstractions. If we are to understand designing then we must understand the origin of the concepts which drive it and the reasoning of the agent. In this paper we describe how such concepts and reasoning by an autonomous and rational agent must be inherently situated.

2. Autonomy and Rationality

An agent is a finite system with a constraint of rationality on its beliefs, goals and actions (Russell & Norvig 1995). The agent has goals and beliefs, and executes actions based on those beliefs and goals. Beliefs are models that the agent constructs over time from experiences by comparing the effects of actions with resulting percepts. Deciding which actions are necessary to achieve goals is task requiring that the agent be rational, but the condition that it be finite means that reasoning is bounded in time and space.

If a design agent is to be rational than it must have beliefs and must undertake some of those actions that are apparently appropriate according to those beliefs (Cherniak 1986). An agent is called autonomous if its beliefs and behaviour are determined from its own experience. This means that the agent's only knowledge of its environment is via actions and percepts. By implication this means that the knowledge of one agent may be different from that of another agent of originally identical construction. The further implication for design agents is that it must design from its own experiences or else not be autonomous. If design knowledge is encoded then the application of that knowledge to a design problem will not be autonomous, and any resulting design could not be solely attributed to the agent.

3. Concepts and Categories

Concepts are abstractions of experience that confer a predictive ability for new situations (Rosenstein & Cohen 1998). They are not static definitions of the members of a class. The concept of a stool is a representation of the activities of the agent with which stools are involved, and its meaning is its predictions of possible interaction. If concepts are abstract then they are intensional (Langley 1996). If that is the case, then what are their extensions ? After all, one cannot have a concept of the class of objects we call \stools" without their being some properties that they share. We shall call categories the extensions of concepts (Cohen 1997).

Consider, for the sake of argument, a human or android that has certain basic concepts like naive beliefs about gravity and solid shapes. For our example we shall consider that the agent has a concept of stool and that it wants to sit on one when available. A concept of stool will be a set of representations of the agents activities with a stool, involving a number of categories:

- visual perceptual categories: what a stool looks like.
- proprioceptive perceptual categories: what it feels like to sit on a stool.
- action categories: the activation of effector signals to sit down and to stand up.
- internal state categories: that sitting feels “comfortable” or not, requires less energy, makes the android feel “happy” and so on.

After lots of interactions with different objects using the above action, the agent will learn what visual categories of \stool shape attended" are suitable and what are not. These will be things like:

- upper “shape” has upper “line” which is at angle “horizontal”
- various cases of the number of “legs”: zero, three evenly distributed, and so on
- height of such-and-such.

As the agent learned abstract concepts like “on” and plurality, these categories would be adjusted. A concept of \table" would employ many similar visual categories to “stool” but the proprioceptive categories would be different, as would the actions involved.

The point of this example is twofold. Firstly, any beliefs that the agent learns about shape will begin from the concrete and extensional and end with the abstract and intensional, not the other way around. Secondly, such beliefs will be situated and embodied. If the agent lives in a world of straight lines it will not have cause to learn categories and concepts which abstract to spheres and the like. The agent starts with simple categories constructed from unabstracted sense-data, and abstract shapes such as \square" are constructed later. We do not believe in shape perception, for instance, as an encoded a priori set of atomic shapes to be later matched against encountered sense-data. Beliefs about shape should be learned just as the consequences of actions are. Initially, visual categories will be quite close to sensory data, but after encountering many objects some of the common properties amongst a category will be abstracted.

If an agent is autonomous then its beliefs must be representations in the first person, that is, situated. Concepts and categories must be abstracted from situated and embodied interactive beliefs. If we view a design agent as autonomous then we cannot escape from the view that all of its beliefs must, in the end, be grounded in interactive experience.

4. Situatedness

Situated agents have beliefs which are first person constructions, not third person encodings. According to a situated view,

> “all action is embodied because perception and action arise together automatically: Learning is inherently “situated” because every new activation is part of an ongoing perception-action coordination” (Clancey 1993).

Situated beliefs are subjective, context dependent, and are grounded in and inseparable from environmental interaction. “Situated” does not mean that models are or are not symbolic, although for the models to be valid and not encodingist any symbols must be grounded in the interactive experiences of the agent (Bickhard & Terveen 1995, Bickhard & Campbell 1996). Learning and interaction are fundamentally important to a situated agent, and memory reflects how it has adapted to its environment. Recall is a synthesising process in which \what must have occurred" is reconstructed from a few fragments in working memory (Cherniak 1986). It is in contrast to the traditional filing cabinet view of memory which is common in AI.

Additionally, by “situation” we do not mean a symbol which references a single description of the current situation. There is no single thing describing the current situation from which a memory access could key. Rather, the interactions and learning of the agent are situated and the “current situation” is something which is distributed amongst all of the modules comprising the agent. Of particular relevance are current goals, biases to sensation and perception processes, and lateral links between percepts of different modalities. For example, it might be that one sensory modality is primarily responsible for the interpretation of a currently attended object. Nevertheless, other modalities continue to operate, influencing each other and conception to lesser extents. The agent as a whole and as an entity interacting with the environment collectively comprise the “situation”.

5. Applying Situatedness to Conceptual Designing

To reiterate, if an agent is to design autonomously then it must apply its learned beliefs as a design evolves. If its beliefs are situated then the application to design problems must also be situated. Now, if an agent designed by interacting with an external representation, then the same perceptual categories and concepts could be invoked as when these categories and concepts were learned from \real" world interaction. Further, by concentrating the processes of perception and sensation onto the current focus of attention after the fashion of deictic reference (Agre 1997) and active vision (Ballard, Mayhoe, Pook & Rao 1997), the reasoning of the agent can remains situated whilst designing.

We have proposed (Smith 2000, Smith & Gero 2000) a design agent which interacts with an external environment which is itself a model of an evolving design. Concrete details of an evolving design reside in the external representation, but these are uninterpreted and mean nothing by themselves. Internal representations comprise a virtual world for the agent that is purely conceptual and exists only in the interpretations of the agent. Internal representations do not necessarily correspond to a world model in the conventional AI sense.

For design the same perceptual categories are used on the external representation as have been learned previously. Actions affecting the external representation will be different, as will be the agent's goals. Also, the perceptual expectations and hypotheses of the agent will be different as during design the agent attempts to determine what a depiction could be. When interacting with a real world it is trying to determine what pre-existing objects there are. An agent, according to our model, makes repeated references to its environment, interpreting and reinterpreting previous and ongoing depictions through constructive processes of sensation and perception.

A storage metaphor of memory would have it that an agent observes objects in its environment, abstracts and then stores models of those objects (Clancey 1991). Later, during perception, a function would be activated repeatedly to match the stored models against the current input. However, this ignores that beliefs are learned through situated interaction. Percepts should be constructed through interaction and beliefs learned through errors in those interactions (Bickhard & Terveen 1995, Bickhard & Campbell 1996), and perceptual beliefs are behaviours of how to perceive. Perception in our view is the agent taking perceptual actions, whose results will vary according to the current situation and the currently attended object, until it is satisfied with its interpretation.

Sensation is a cellular and spatially distributed process of feature instantiation out of raw input data. These input cells could be pixels from a camera or graphics buffer, or could be receptive fields in humans. For each sensory modality, sensation performs a representation shift from the space of raw input into something more appropriate for learning and reasoning. As an example, colour might be discretised from a continuous colour space into a histogram space. Sensation is also biased by perception. Taking the colour example again, biases might include brightness, contrast, discretising parameters and other contiguous colours.

Sense-data are combined by processes of perception into the categories that are used both to interpret further interactions and as the extensions of concepts. An agent performs an overt depiction action, looks to the external representation to interpret the depiction and decides what action to perform next. This looking requires finding an object in a depiction (localisation) as well as interpreting what is at the current focus of attention (identification). Perception, then, is an interactive process, a “rapid sequence of microperceptions and microactions” (Damasio & Damasio 1992) occurring amongst the various sensory modalities.

Percepts are constructed based on perceptual actions taken on current sense-data, being biased by expectations generated both at the perception level and from concept-driven hypotheses. The percepts provide what is necessary to discriminate, through interaction, between possible interpretations of the currently attended object. They trigger conception to interrogate the environment, and environmental feedback influences the course of further perceptual interrogations (Thomas 1999). The agent makes repeated references to its environment, interpreting and reinterpreting previous and ongoing depictions through constructive processes of sensation and perception. The process of perception is as important in this model as the resulting percepts, and is based on scanning the focus of attention around a localised object to interpret shape.

6. Discussion

We do not believe that the problem space search model of conceptual design is sufficient computationally. If sufficient a priori knowledge existed in the agent to allow internal inference of an object or relation, P, in bounded time and space then the discovery of P would not properly be characterised as a conceptual design task. If it could succeed but only by making guesses then such reasoning would not be a rational.

Problems like encodingism (Bickhard & Terveen 1995, Bickhard & Campbell 1996) and the nature of autonomous learning lead to the conclusion that beliefs must be situated and learned in the first person through interactive experience. Further, consideration of the advantages of external representations both computationally and psychologically (Sloman 1995, Zhang 1997, Kulpa 1994, Koedinger 1992, Hegarty 1999) suggest a mechanism for situated beliefs of an agent to be applied during design. It is our belief that a situated metaphor, one based on interaction with an external representation, is more appropriate for conceptual designing than a problem space search metaphor.

References

Agre, P. E. (1997). *Computation and Human Experience*, Cambridge University Press, Cambridge.

Ballard, D. H., Mayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997). Diectic codes for the embodiment of cognition, *Behavioral and Brain Sciences* **20**(4): 723-742.

Bickhard, M. H. & Campbell, R. L. (1996). Topologies of learning, *New Ideas in Psychology* **14**(2): 111-156.

Bickhard, M. H. & Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution,* North-Holland Elsevier Science, Amsterdam.

Cherniak, C. (1986). *Minimal Rationality,* MIT Press, Cambridge MA.

Clancey, W. J. (1991). Book review: Israel Rosenfeld, The Invention of Memory: A New View of the Brain, *Artificial Intelligence* **50**(2): 241-284.

Clancey, W. J. (1993). Situated action: A neurobiological interpretation response to Vera and Simon, *Cognitive Science* **17**(1): 87-116.

Cohen, P. R. (1997). Acquiring conceptual structure through activity, *Seminar*, Carnegie Mellon University. *http://eksl-www.cs.umass.edu/research/conceptual-systems/

Damasio, A. R. & Damasio, H. (1992). Brain and language, *Scientific American* pp. 63-71.

Gero, J. S. (1998). Conceptual designing as a sequence of situated acts, *in* I. Smith (ed.), *Artificial Intelligence in Structural Engineering,* Springer, Berlin, pp. 165-177.

Hegarty, M. (1999). Mental animation, *Seminar*, Key Centre for Design Computing and Cognition, University of Sydney.

Koedinger, K. R. (1992). Emergent properties and structural constraints: Advantages of diagrammatic representations for reasoning and learning, *Proc. of AAAI-92,* http://act.psy.cmu.edu/ACT/papers/AAAIT92.html

Kulpa, Z. (1994). Diagrammatic representation and reasoning, *Machine Graphics and Vision* **3**(1): 77-103.

Langley, P. (1996). Elements of Machine Learning, Morgan Kaufmann Publishers, San Francisco.

Rosenstein, M. T. & Cohen, P. R. (1998). Concepts from time series, *Proc. of Fifteenth National Conference on Artificial Intelligence,* AAAI, Menlo Park, CA.

Russell, S. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach,* Prentice Hall, Englewood Cliffs, NJ.

Sloman, A. (1995). Musings on the roles of logical and non-logical representations in intelligence, *in* J. Glasgow, N. N. Narayanan & B. Chandrasekaran (eds), *Diagrammatic Reasoning: Cognitive and Computational Perspectives,* MIT Press, Cambridge, MA, pp. 7-32.

Smith, G. J. (2000). Conceptual design through interaction with an external representation. *PhD Thesis Proposal,* Department of Architectural and Design Science, University of Sydney.

Smith, G. J. & Gero, J. S. (2000). Conceptual design through interaction with an external representation, *Poster Abstracts,* Artificial Intelligence in Design'00. (to appear)

Thomas, N. J. T. (1999). Are theories of imagery theories of imagination ? An active perception approach to conscious mental content, *Cognitive Science* **23**(2): 207-245.

Zhang, J. (1997). The nature of external representations in problem solving, *Cognitive Science* **21**(2): 179-217.

This paper is a copy of: Smith, G. and Gero, J. S. (2000) The autonomous, rational design agent, *in* H. Fujii (ed.) *Workshop on Situatedness in Design*, Artificial Intelligence in Design'00, Worcester, MA, pp. 19-23.