Commitment and Effectiveness of Situated Agents

April, 1991

Technical Note 17

By:

David N. Kinny
Department of Computer Science
University of Melbourne
Parkville 3052, Australia

Michael P. Georgeff
Australian Artificial Intelligence Institute
1 Grattan Street
Carlton 3053, Australia

This paper is to appear in the *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence, IJCAI '91*, Sydney, published by Morgan Kaufmann Publishers, Mountain View, California, 0990. It is also available as Technical Report 90 0 , Department of Computer Science, The University of Melbourne.

Abstract

Recent research in real-time Artificial Intelligence has focussed upon the design of situated agents and, in particular, how to achieve effective and robust behaviour with limited computational resources. A range of architectures and design principles has been proposed to solve this problem. This has led to the development of simulated worlds that can serve as testbeds in which the effectiveness of different agents can be evaluated. We report here an experimental program that aimed to investigate how commitment to goals contributes to effective behaviour and to compare the properties of different strategies for reacting to change. Our results demonstrate the feasibility of developing systems for empirical measurement of agent performance that are stable, sensitive, and capable of revealing the effect of "high-level" agent characteristics such as commitment. Such systems are likely to have an increasing role to play in guiding the design of situated agents for specific domains, and in contributing to a better understanding of how the characteristics of agents and environments interact.

1 Introduction

The crucial problem facing designers of *situated agents* – artificial systems capable of effective, rational behaviour in dynamic and unpredictable environments – is to ensure that the agent's responses to important changes in its environment are both appropriate and timely. These requirements appear to conflict, since the reasoning that seems to be needed to choose appropriate actions could require an arbitrarily large amount of time to perform. Somehow this reasoning must be limited and controlled – there must be a *rational balance* between reasoning and acting.

Designs and architectures that address this problem can be placed on a spectrum according to the amount of reasoning they perform. At one end are reactive systems [Agre and Chapman, 0876, Schoppers, 0876] that minimize the need for run-time computation by precompiling appropriate responses to situations into a form that can be utilized without reasoning. At the other extreme are *real-time reasoning systems* [Georgeff and Ingrand, 0878, Fehling and Wilber, 0878] that attempt to achieve rational balance by reasoning based upon explicit representations of their beliefs, goals, and intentions. Purely reactive systems are capable of effective behaviour in some dynamic environments, but cannot guarantee it in the unanticipated situations that are inevitable in real-world domains. Real-time reasoning systems can behave more robustly in such situations because of their ability to reason about multiple conflicting goals, how best to achieve them, their relative value and urgency, etc.

Recent theoretical work has clarified the role of goals, intentions, and commitment in constraining the reasoning that an agent performs [Cohen and Levesque, 088 , Rao and Georgeff, 0880] and has examined the trade-off between reaction and deliberation [Bratman *et al.*, 0877]. However, these theories are quite general, and say little about specific real-time reasoning strategies and their effect on agent behaviour. To evaluate the effect of such strategies, we require (in the absence of a comprehensive theory) a controlled environment in which we can conduct experiments, and appropriate measures of agent performance so that different strategies can be meaningfully compared. Such an environment, called a *simulated world*, should ideally capture the essential features of real-world domains while permitting flexible, accurate, and reproducible control of the world's characteristics. A number of such worlds have been developed [Cohen *et al.*, 0878, Pollack and Ringuette, 088], but there have been few published reports of the results of experimental evaluation of agent performance.

This paper describes the experimental work we have been doing in this direction. The main aims of the work were to9

- Assess the feasibility of experimentally measuring agent effectiveness in a simulated environment.
- Investigate how commitment to goals contributes to effective agent behaviour.
- Compare the properties of different strategies for reacting to change.

The experimental system was based upon the PRS real-time reasoning system [Georgeff and Ingrand, 0878] operating within the Tileworld environment [Pollack and Ringuette, 088]. In the next section we briefly review the essential features of the experimental system and introduce the terminology used in our analysis. We then present and analyze the results of several sequences of experiments.

2 The Experimental System

2.1 The Environment

The Tileworld [Pollack and Ringuette, 088] is a comprehensive testbed for experimental evaluation of agent performance. The number of independent control parameters is large, and the problems an agent needs to solve in order to function effectively in this domain are non-trivial. However, such a testbed was considered too rich for the initial investigative experiments we had planned. Therefore, to reduce the complexity of the object-level reasoning required of our agent, we employed a simplified Tileworld with no tiles.

In essence, our Tileworld is a 1-dimensional grid on which an agent scores points by moving to targets, known as *holes*. When the agent reaches a hole, the hole is *filled*, and disappears. The task is complicated by the presence of fixed obstacles. The Tileworld is 3-connected: the agent can move horizontally or vertically, but not diagonally. A lower bound on the shortest path length between two points is thus given by the the Manhattan distance $(\Delta_x + \Delta_y)$.

Holes appear in randomly selected empty squares, and exist for a length of time known as their *life-expectancy*, unless they disappear prematurely due to the agent's actions. The actual time for which a hole exists is its *lifetime*. The interval between the appearance of one hole and the next is known as the *gestation period*. Each hole has a specific value, its *score*. Life-expectancies, gestation periods, and scores are taken from independent random distributions.

2.2 The Agent

We chose to use the PRS real-time reasoning system [Georgeff and Ingrand, 0878] to construct a simple but adequately competent agent in which the effects of different object-level and meta-level strategies could be quantified. Accordingly, we made a number of simplifying assumptions in its design9

- The agent has perfect, zero-cost knowledge of the current state of the world, but no knowledge of its future. We ignored issues of uncertainty.
- The agent forms only correct and complete plans. We did not attempt to explore partial planning, anytime algorithms [Dean and Boddy, 0877], etc.
- The agent plans only at a single level, forming plans consisting of a sequence of steps to a single hole, rather than multiple hole tours.

Our agent is based upon a simple plan act cycle. In the planning phase, it uses a path planner to produce a (shortest path) plan for each current option (hole), and a *plan selection strategy* to select one of these plans. In the action phase, part or all of the chosen plan is executed. Various attributes of this process are parameterized, allowing a range of different agent behaviours to be produced. The time cost of planning, known as the *planning time*, is also a controlled parameter. (The actual computational cost of planning is immaterial.)

Plan selection strategies choose between plans on the basis of the distance, score, and age of the target hole. A utility function applied to these attributes produces an integer-valued measure of the value of the plan, and the plan with the maximum value is selected. We implemented a number of utility functions that selected among plans on the basis of different weightings of these attributes.

2.3 Degree of Commitment and Reactivity

One of the primary aims of our work was to investigate different aspects of agent commitment and its effect on performance. First, we wished to examine the effect of an agent's degree of commitment to its current plan, independent of events in the world. Pollack and Ringuette [088] define a *bold* agent as one that never reconsiders its options before the current plan is executed in its entirety, and a *cautious* agent as one that reconsiders every new option. We use the same terminology here, but in a restricted sense.

Our agent is potentially faced with a new set of options after each step it takes, both because the world changes and because its plan utility function can be position sensitive. We implemented a crude parameterization of its degree of boldness by specifying the maximum number of plan steps that the agent executes before replanning. A value of 0 produces a cautious agent that replans after every step, while a sufficiently large value produces a bold agent.

Second, we wished to investigate how sensitivity to change could affect agent behaviour. Such change could potentially trigger replanning, but in dynamic domains replanning must be limited or action will not be sufficiently timely. The key is to decide cheaply using some simple estimate of utility whether an event warrants more expensive deliberation [Russell and Wefald, 0878]. PRS is particularly well-suited for the implementation of such *event filters* [Pollack and Ringuette, 088] because of its reactive meta-level processing capability. By appropriately specifying the *meta-level invocation criteria* [Georgeff and Ingrand, 0878], a range of sensitivities to external change can be built into the agent's reflective strategies.

In this way, we implemented several different *reaction strategies* to control what events would trigger replanning, and so explored a number of types of *rational* commitment, as opposed to the *blind* commitment of an agent that continues to execute its plans without regard to external events.

2.4 Performance Measurement

The Tileworld and our agent both measure time by an abstract clock. They execute synchronously, with the ratio of their clock rates set by a parameter γ called the *rate of world change*. This parameter allows the dynamism of the world, as perceived by the agent, to be varied over a wide range.

The natural notion of how well an agent performs in the Tileworld is its score, or rather the sum of the scores of the holes it has filled. Absolute score is of little use as a measure of effectiveness, as hole scores are taken from random distributions whose characteristics are parameterized, experiment lengths vary, and worlds can have different characteristics and speeds.

Previous experimental results [Pollack and Ringuette, 088] had revealed the undesirability of using performance measures based on CPU-time or elapsed time, and our abstract time unit was arbitrary. Accordingly, we defined an agent's effectiveness $\mathcal{E}$ to be its score divided by the maximum possible score[1] it could have achieved. This gave a measure of performance that was largely independent of game length, and proved to be stable and reproducible.

[1] The sum of the scores of all holes appearing in the trial.

2.5 Experimental Procedures

During the course of a single Tileworld game, an agent's effectiveness fluctuates, due to random variations in hole scores and positions and the fact that the agent scores by increments rather than continuously. An individual game had to be sufficiently long for the agent's effectiveness to converge to a stable value. By plotting the effectiveness $\mathcal{E}$ as a function of game length, we could determine what this length should be.

Across different, but statistically similar games, we observed small variations in agent effectiveness whose magnitude declined as game length increased. This led us to define a *standard experiment* consisting of 4 games with identical initial configurations, but differing in the random seed that determined their evolution. The variation in effectiveness was recorded and served as a confidence limit on the mean effectiveness value. Each such experiment produced one point on a graph. A *characterization* of an agent consisted in running such an experiment at up to 04 points spanning a 1^5 range of values in γ. This produced a curve such as that shown in Figure 0. The error bars indicate the measured variation in $\mathcal{E}$. We will usually omit them for clarity. Their magnitude is typical of our standard experiments.

The stability of the results of such standard experiments was excellent. Each required about 2 minutes of CPU-time on a Sun Sparcstation, with a characterization taking up to 7 hours.

3 Results and Analysis

In this section we present the results of several sequences of experiments aimed at investigating how agent effectiveness changes as some world or agent parameter is varied. The parameters selected for investigation were9

- Rate of world change
- Agent planning time
- Degree of commitment
- Reaction strategy

The results that follow all use a plan selection strategy that maximizes hole score divided by distance, which is arguably "rational" since this is a measure of the rate at which the agent can hope to score. It is similar to the *subjective expected utility* of Pollack and Ringuette [088]. Experiments comparing different plan selection strategies are reported elsewhere [Kinny, 088].

3.1 Rate of World Change

The experimental parameter that has the most fundamental influence upon agent effectiveness is the rate γ at which the world changes. Figure 0 shows the effectiveness $\mathcal{E}$ of a bold agent as a function of γ across 1 orders of magnitude.

At the baseline rate of world change ($\gamma = 0$) in our standard world, holes appear and disappear sufficiently slowly that any agent with a moderately effective strategy will successfully fill each hole soon after it appears and achieve a perfect score ($\mathcal{E} = 0$).

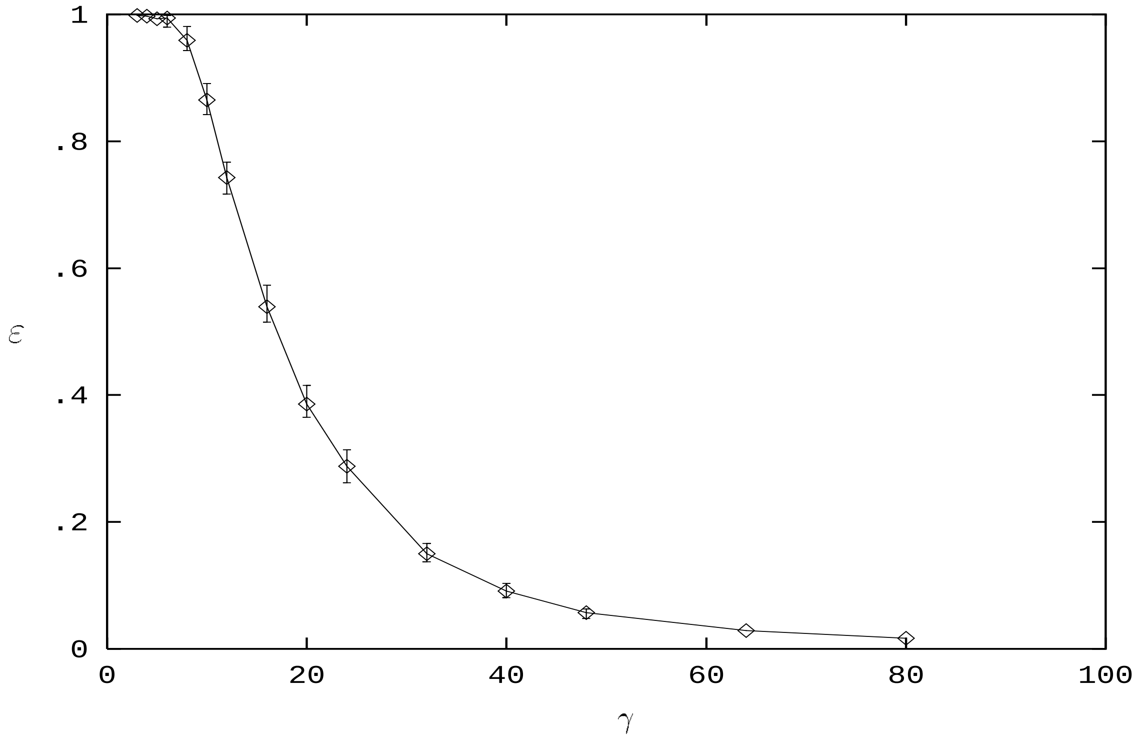

Figure 09 Effect of Rate of World Change

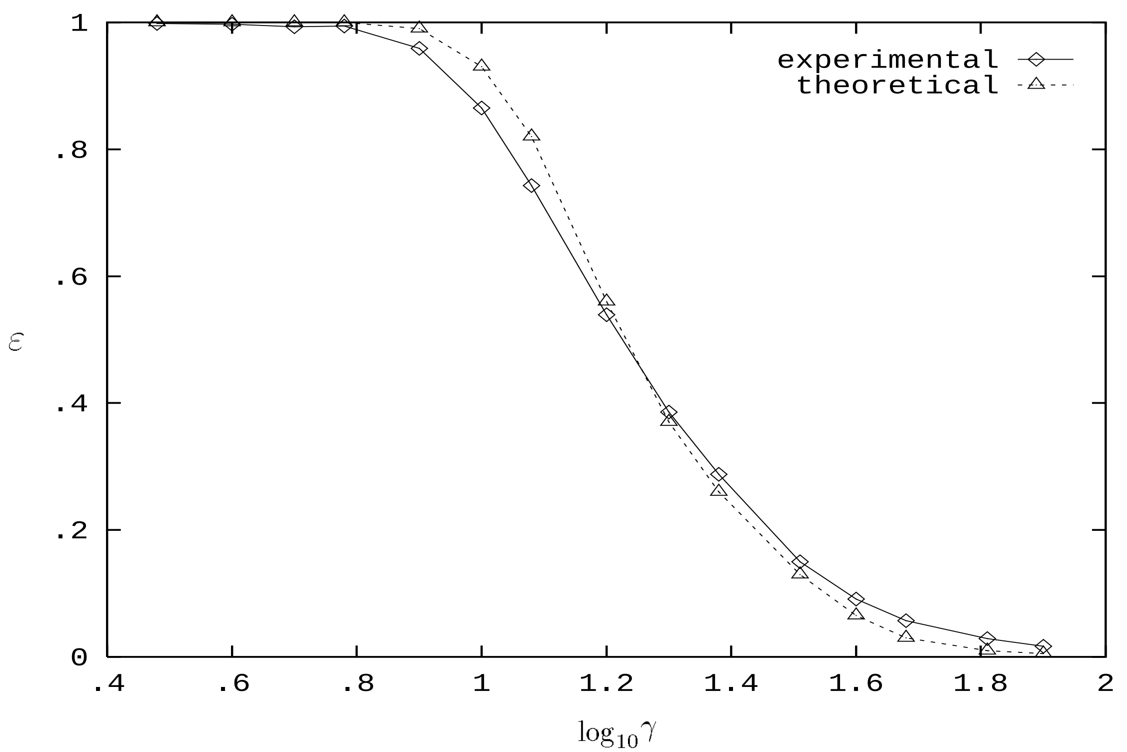

Figure 19 Effect of Rate of World Change (log x-scale)

As γ increases, hole life-expectancies decrease reciprocally, as measured by the agent's clock. Eventually holes start to disappear before the agent has filled them, and ε drops below 0. This decline in effectiveness has a sudden onset and is initially steep, with ε falling from .8 to .4 for a factor of 1 increase in γ. As γ increases further the decline in ε becomes more gradual and eventually asymptotically approaches zero.

The experimental performance curve in Figure 1 shows the same data plotted against $\log_{10}\gamma$. The log scale spreads the decline in ε more uniformly, making comparison of curves easier. We refer to such a plot of ε vs $\log_{10}\gamma$ as an effectiveness or performance curve. Families of such curves provide a means for comparing the effect of change in other experimental parameters, and are the standard method of presenting such results in subsequent sections.

An explanation of the shape of these effectiveness curves can be obtained by considering the equilibrium between hole appearance and the agent's hole filling activities. A given hole's life-expectancy l and gestation period g are taken from independent uniform distributions. At the baseline rate of world change, these are9

Parameter	Minimum	Average	Maximum
Gestation g	5	04	13
Life-expectancy l	13	5	85

Were it not for the activities of the agent, a hole's lifetime would be equal to its life-expectancy, and the expected number of holes in the world h_{ave} would be given by $h_{\text{ave}} = l_{\text{ave}}/g_{\text{ave}} = 3$.

The total time f the agent takes to fill a hole is determined by the time spent planning, moving, and replanning. For an agent that reconsiders its options every k steps, this is given by $f = d(p/k + m)$, where p is the planning time, d is the distance to the hole (or more precisely the path length of the plan), and m is the time to move a single step (henceforth always 0). Setting $k = d$ gives a bold agent that commits to executing its entire plan, while setting $k = 0$ gives a cautious agent that replans after every step. For the curve in Figure 1, the average filling time f_{ave} is approximately 04.

When this time is smaller than the minimum hole gestation period g_{min}, as is the case when $\gamma < 3$, then h_{ave} will be less than 0. The agent will spend only a fraction of its time actually filling holes or planning: the rest of the time it will be waiting for a hole to appear. As γ increases this idle time will decrease to zero, and h_{ave} will rise above 0.

For $g_{\text{min}} < f < 1g_{\text{min}}$, usually at most one new hole will appear while the agent is filling a prior one, so despite the fact that $h_{\text{ave}} > 0$, a bold agent rarely has to choose which hole to fill next. Note, however, that often the *task timelimit* (the time left for the agent to fill a selected hole) will be less than the hole's life-expectancy, because of the delay between a hole's appearance and the agent's targeting it. For $1g_{\text{min}} < f < g_{\text{ave}}$, the agent will more often have to choose between possible targets, but nonetheless will still be able, on average, to fill holes faster than they appear. Sometimes it will miss a hole, so ε begins to drop below 0. At this point, the agent's plan selection strategy becomes important.

When $f > g_{\text{ave}}$, holes are appearing faster than the agent can fill them, so necessarily the population h_{ave} will rise. The average time between a hole appearing and the agent trying to fill it will increase, leading to shorter task timelimits. Thus the probability that a hole will disappear before being filled will increase and ε will decline. To begin with, this decline will be rapid. Ultimately, the rise in h_{ave} will be limited by the upper bound of the holes' natural life-expectancies l and the decline in ε becomes less steep. The agent succeeds in

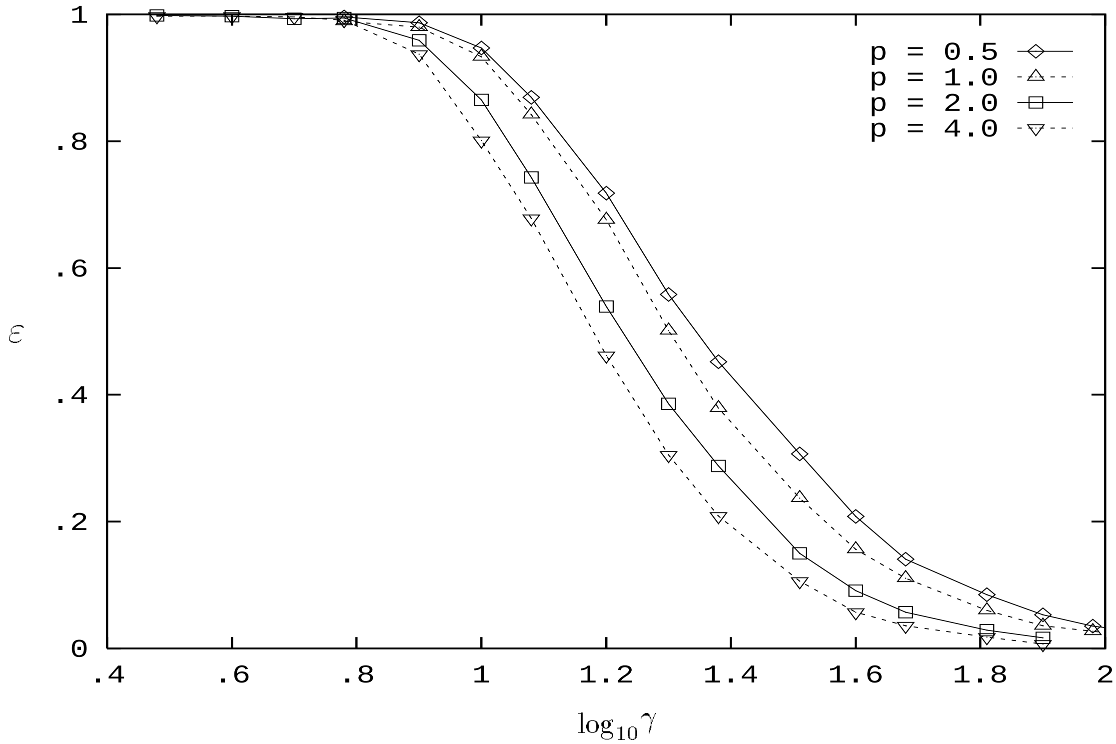

Figure 29 Effect of Planning Time (bold agent)

filling those holes whose life-expectancies are sufficiently long or that are targeted sufficiently promptly. Even in highly dynamic worlds, there will be some holes that meet these criteria – hence the asymptotic behaviour. Note that the agent may successfully fill only some small fraction of the holes, but by judicious choice of targets may achieve a significantly higher value of ε.

We can put this informal explanation of the shape of an agent's performance curve onto a mathematical footing by calculating the distribution of path lengths in this domain. Ignoring the effect of the agent's plan selection strategy on the distribution of selected path lengths, we can approximate ε by finding what fraction of these paths result in f being smaller than the average task timelimit, which can be calculated[2] from the average hole population and life-expectancy. This leads to the theoretical performance curve in Figure 1. The agreement between the curves is good, the noticeable difference being that the decline in the experimental curve begins earlier but is less steep. This is partly due to our ignoring the fact that the hole life-expectancies come from a distribution.

3.2 Agent Planning Time

The planning time p is the agent parameter that determines the cost of forming a plan. If it is small, the agent can afford to replan often. If it is large, frequent replanning is not effective.

Figure 2 shows the performance curves for a bold agent with p equal to .4, 0, 1, and 3. These performance curves are near-identical, differing only by an x-axis offset. Since a given level of effectiveness is associated with a given ratio of hole life-expectancy l to filling time

[2] For details see [Kinny, 0990]

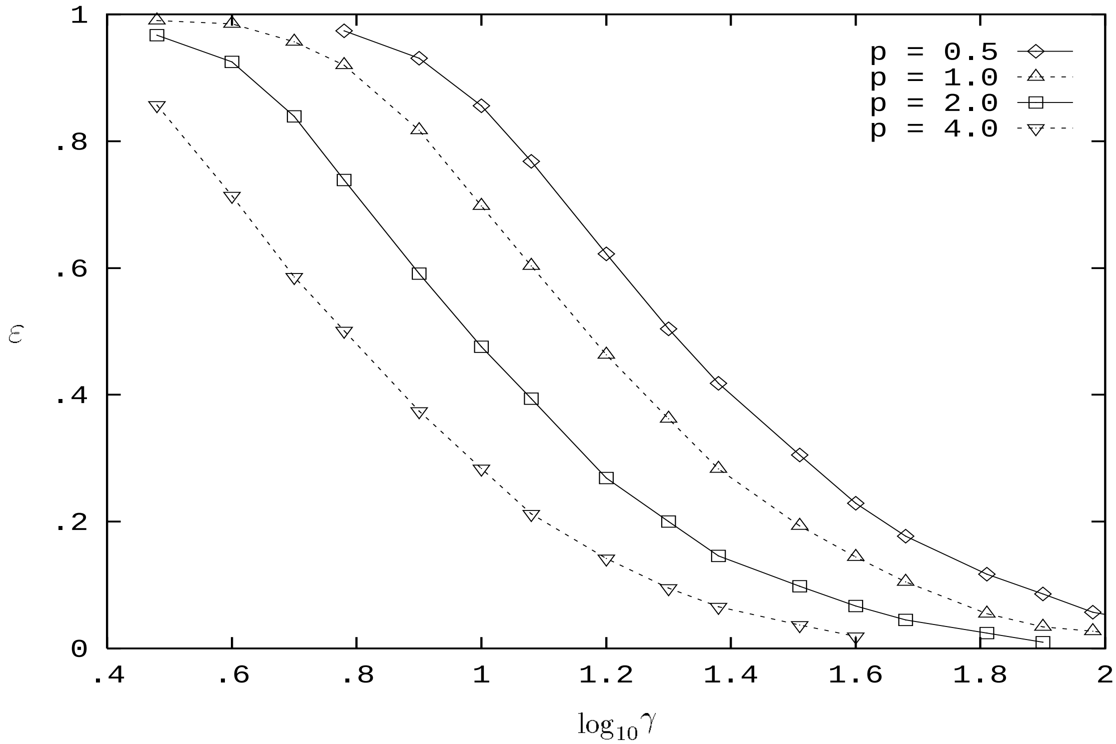

Figure 39 Effect of Planning Time (cautious agent)

f, and l $0/\gamma$, the effect on ε of varying p, and hence multiplying f by some constant c, is equivalent to multiplying γ by c, i.e. scaling the x-axis. Since the x-axis is a log scale this appears as a constant offset.

Figure 3 shows the performance curves for a cautious agent with p equal to .4, 0, 1, and 3. Again we observe a family of similar curves differing by a scale factor, but the falloff in performance as p rises is far more pronounced, with the curves being more widely separated. The filling time f is more sensitive to p, since replanning occurs after every step. The other noticeable difference is the reduced rate of decline in ε as γ increases. This is attributable to the fact that, all other things being equal, f is higher for a cautious agent than a bold agent, hence the point at which ε begins to decline occurs earlier. The decline is slower, however, because the proportion of f contributed by p is independent of distance, whereas for a bold agent it increases for shorter path lengths. Another contributing factor is the ability of the cautious agent to be opportunistic. This is discussed in the next section.

3.3 Degree of Boldness

We have seen differences between bold and cautious agents as planning time is varied. In this section we consider in more detail the effect of boldness on performance, also examining agents that are intermediate in boldness in that they commit to executing several steps of a plan before reconsidering their options.

Figure 4 shows the performance curves of 2 agents differing only in the length of the plans to which they commit. The cautious agent replans after 0 step, the "normal" agent after 3, and the bold agent after the entire plan has been executed. All have planning time $p = 3$. In this case the bold agent is superior to the normal agent, which is superior to the cautious

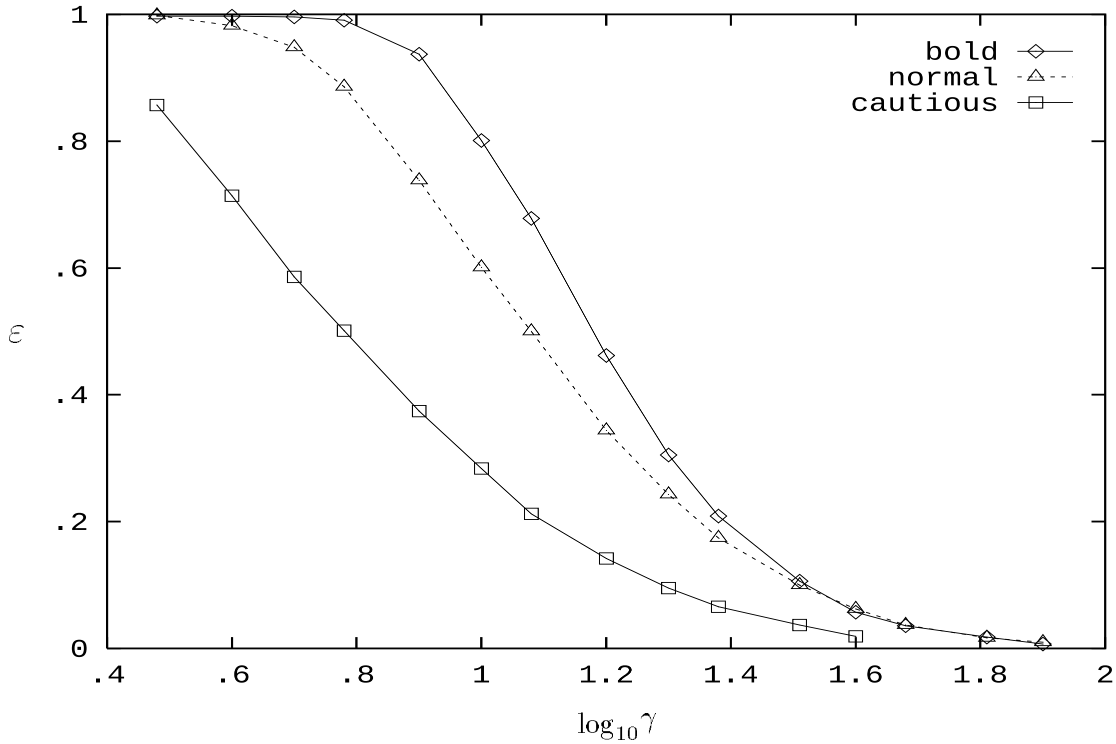

Figure 49 Effect of Degree of Boldness ($p = 3$)

agent. The performance curves exhibit the differing slopes that were noted in section 2.1, the normal agent being intermediate between the other two.

An interesting phenomenon appears as we continue to make this comparison while decreasing the planning time p (Figures 5 and 6). As p decreases two things happen9 the gap between the cautious and the bolder agents becomes narrower, and the normal agent becomes superior to the bold agent in more dynamic worlds. In the final graph the cautious agent is as effective as the bold agent at high rates of change.

We can explain this behaviour fairly naturally by considering the tradeoff between the costs and benefits of replanning. In a world where things change slowly, frequent replanning is not advantageous. On the contrary, the increased time spent planning may be an onerous burden. As the degree of dynamism increases, however, boldness becomes a less successful strategy, as it continues to pursue vanished targets or fails to notice better opportunities that have arisen. More cautious agents can be opportunistic and take advantage of serendipitous change, as well as dropping plans that have become futile. In highly dynamic worlds only a small fraction of the agent's plans are successful, and the advantages of opportunism and plan execution monitoring outweigh those of commitment, provided that p is not too great.

3.4 Reaction Strategies

Until now, all committed agents that we have been considering have been blind in their commitment, in that they ignore changes in the world that occur during plan execution. As we saw above, this brings disadvantages as the dynamism of the world increases. Different strategies for reacting to change in the world can improve upon this blind or fanatical commitment.

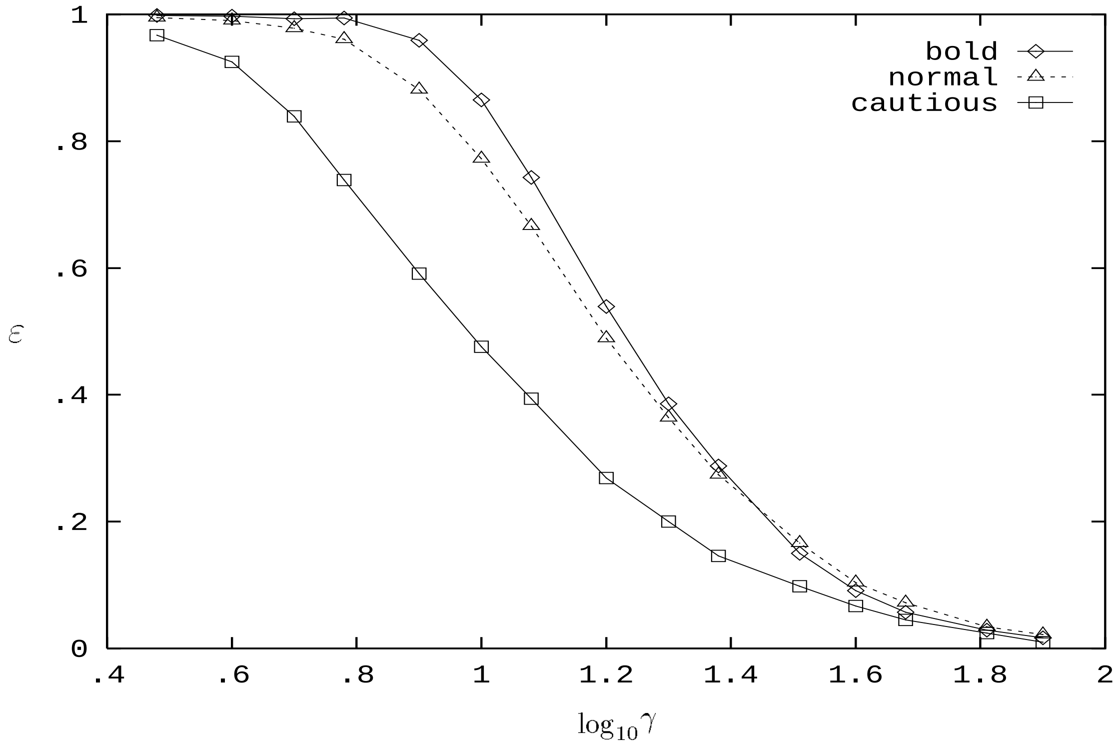

Figure 59 Effect of Degree of Boldness ($p = 1$)

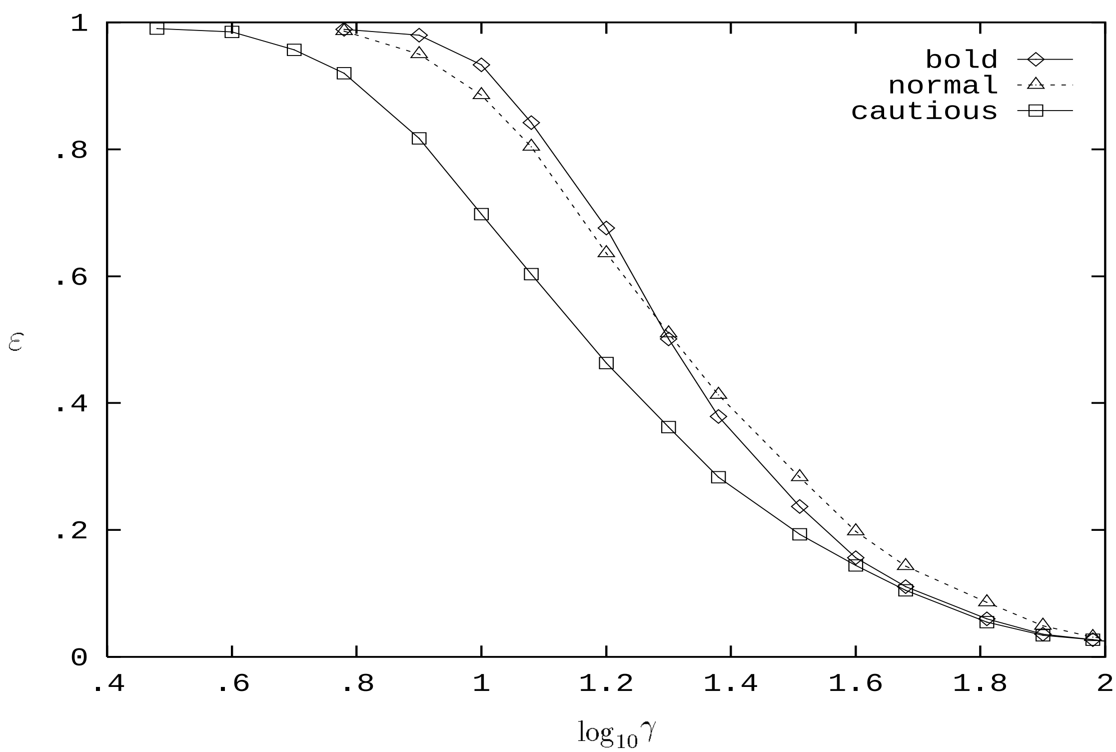

Figure 69 Effect of Degree of Boldness ($p = 0$)

The appropriate response to an event is sometimes to ignore it and other times to deliberate, depending on the agent's current plans, the nature of the event, and the state of the world. The key to effective control of deliberation is to have a reaction strategy that is sensitive to the right environmental cues, producing rational commitment.

We implemented and assessed a range of reaction strategies, including replanning

- when the target disappears,
- when the target disappears or any hole appears, and
- when the target disappears or a nearer hole appears.

The performance curves for these strategies, for a bold agent with $p = 1$, appear in Figure 7. We observe that reacting to the disappearance of the target improves performance significantly. Combining this with replanning when a nearer hole appears is better still. On the other hand, reacting to any new hole is worse than blind commitment, except for high values of γ.

In Figure 8 we see a similar set of results for a bold agent with $p = 0$. As may have been anticipated, the strategy of reacting to every new hole has improved relative to blind commitment, due to the decreased cost of planning.

In the previous section we saw that a more cautious agent was superior to a bold agent in highly dynamic worlds. We investigated whether this still held for more rational types of commitment. In Figure 0 we see the effect of degree of boldness for an agent that replans whenever its target disappears. Comparing these results with Figure 6, it is seen that the simple change in commitment from blind to reactive results in the bold agent being everywhere superior to the more cautious ones.

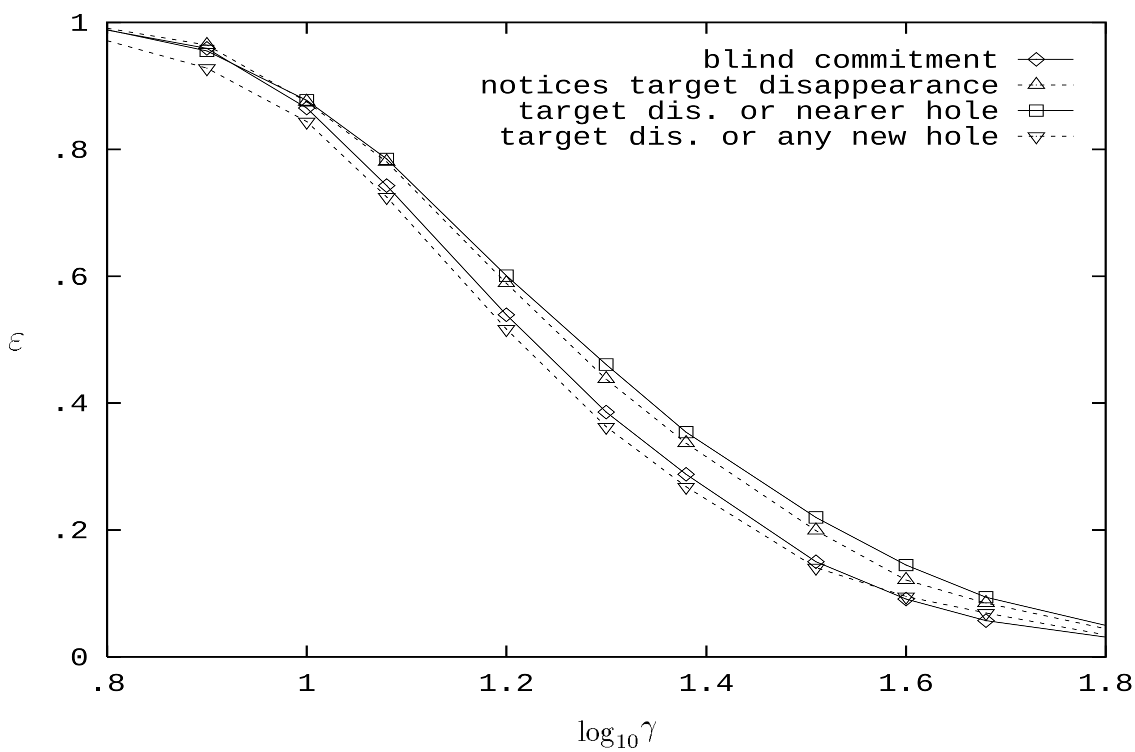

Figure 79 Effect of Reaction Strategy ($p = 1$)

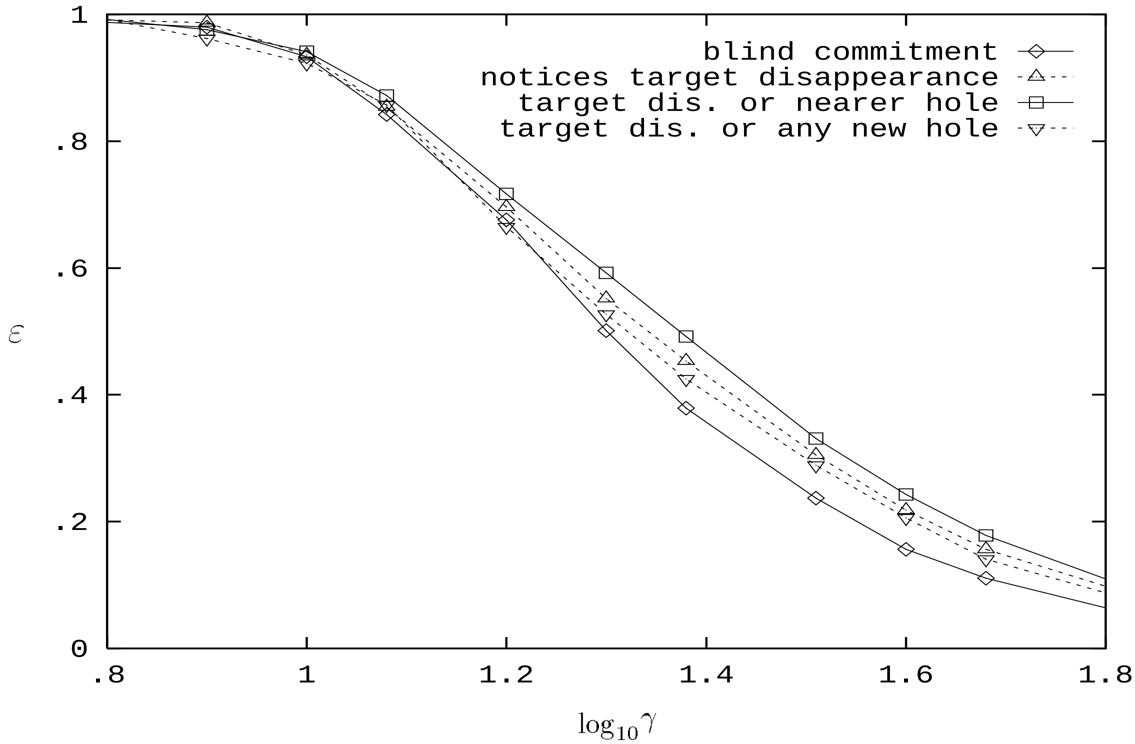

Figure 89 Effect of Reaction Strategy ($p = 0$)

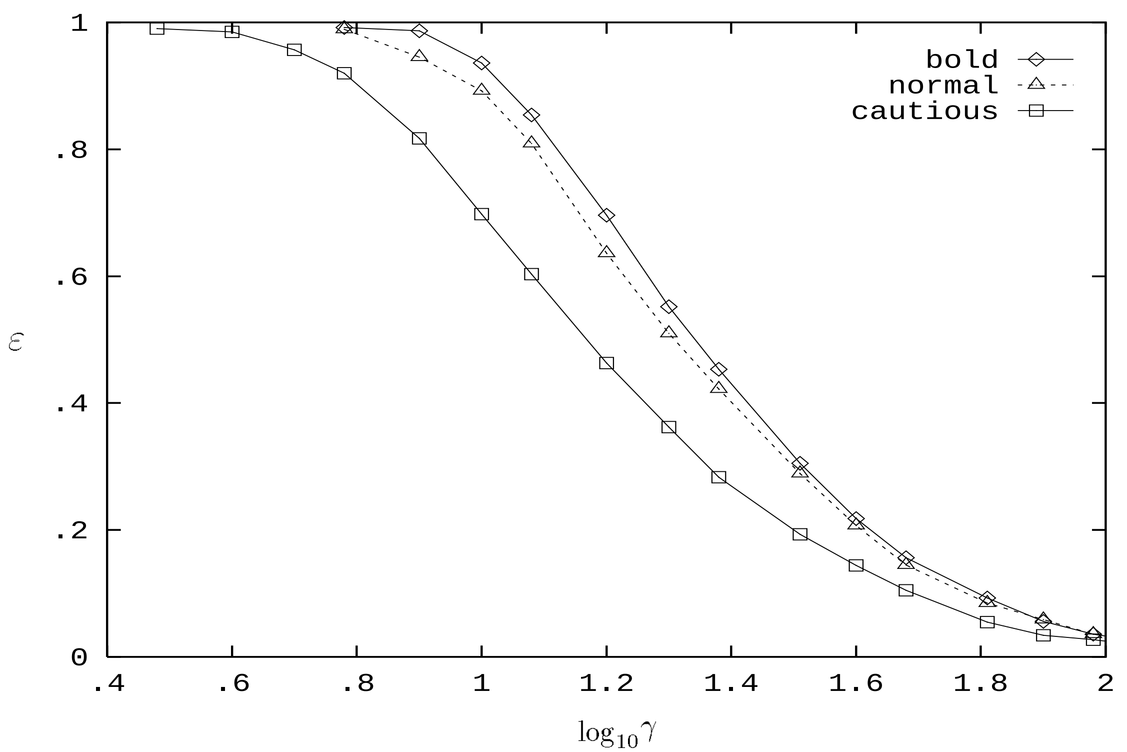

Figure 0 9 Effect of Degree of Boldness (reactive agent, $p = 0$)

4 Conclusions

By combining the PRS real-time reasoning system and the Tileworld simulated environment we have been able to construct a highly parameterized class of agents and environments. The combined system allows the dynamic characteristics of the agent and environment to be varied over a wide range and enables characterization and comparison over a large space of agent environment combinations. We have investigated a small part of that space.

The use of simulated environments to compare the performance of agents and architectures is a technique still in its infancy. However, we have been able to show that, with the appropriate system design and choice of measurement techniques, empirical measurement of agent performance can be stable and sensitive, capable of revealing subtle differences arising from small variations in the agent's control parameters and strategies. Changes in "high-level" characteristics such as commitment have clearly visible effects. These experimental systems are likely to have an increasing role to play in guiding the design of situated agents for specific domains, and in contributing to a better understanding of how the characteristics of agents and environments interact.

Despite the simple nature of our agent's planning and control mechanisms, our results underline the importance of reactive meta-level control of deliberation for resource-bounded agents situated in dynamic domains. For the domains and agents we have explored, the combination of commitment with intelligent reactive replanning was observed to result in optimal behaviour.

Acknowledgements

The authors would like to thank Magnus Ljungberg, Anand Rao, Grahame Smith and Liz Sonenberg for their helpful discussions during the experimental program and for reviewing drafts of this paper. This research was in part supported by a *Generic Industry Research and Development Grant* from the Department of Industry, Technology and Commerce, Australia and in part by the Australian Civil Aviation Authority.

References

[Agre and Chapman, 0876] Philip E. Agre and David Chapman. Pengi9 An implementation of a theory of activity. In *Proceedings of the Sixth National Conference on Artificial Intelligence, AAAI-87*, pages 157–161, Seattle, WA, 0876.

[Bratman *et al.*, 0877] Michael E. Bratman, David J. Israel, and Martha E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 3(3)9238–244, 0877.

[Cohen and Levesque, 088] Paul R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 31(2)9102–150, 088 .

[Cohen *et al.*, 0878] Paul R. Cohen, Michael L. Greenberg, David M. Hart, and Adele E. Howe. Real-time problem solving in the phoenix environment. In *Proceedings of the Workshop on Real-Time Artificial Intelligence Problems at IJCAI-89*, 0878.

[Dean and Boddy, 0877] Thomas Dean and Mark Boddy. An analysis of time-dependent planning. In *Proceedings of the Seventh National Conference on Artificial Intelligence, AAAI-88*, pages 38–43, Saint Paul, MA, 0877.

[Fehling and Wilber, 0878] M. R. Fehling and B. M. Wilber. Schemer-II9 An architecture for reflective, resource-bounded problem solving. Technical Report 726-78-2 , Rockwell International Science Center, Palo Alto Laboratory, Palo Alto, CA, 0878.

[Georgeff and Ingrand, 0878] Michael P. Georgeff and Felix Ingrand. Decision-making in an embedded reasoning system. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, IJCAI-89*, pages 861–867, Detroit, MI, 0878.

[Kinny, 088] David N. Kinny. Measuring the effectiveness of situated agents. Technical Report 00, Australian AI Institute, Carlton, Australia, 088 .

[Kinny, 0880] David N. Kinny. Achieving rational balance. Technical Report 08, Australian AI Institute, Carlton, Australia, 0880.

[Pollack and Ringuette, 088] Martha E. Pollack and Marc Ringuette. Introducing the Tileworld9 Experimentally evaluating agent architectures. In *Proceedings of the Eighth National Conference on Artificial Intelligence, AAAI-90*, pages 072–078, Boston, MA, 088 .

[Rao and Georgeff, 0880] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, Boston, MA, 0880.

[Russell and Wefald, 0878] Stuart Russell and Eric Wefald. Principles of metareasoning. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 3 –300, Toronto, 0878.

[Schoppers, 0876] Marcel J. Schoppers. Universal plans for reactive robots in unpredictable environments. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence, IJCAI-87*, pages 741–748, Milan, Italy, 0876.