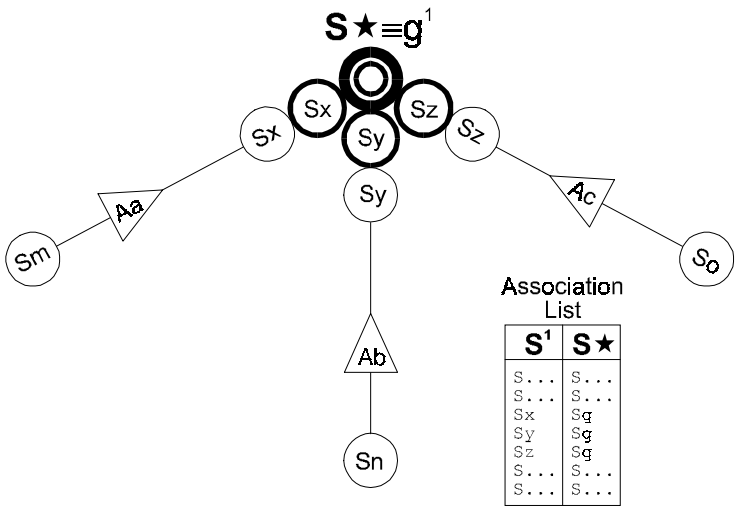
Chapter 7

7. Extensions to SRS/E and Further Work

SRS/E is an experimental system. By their nature such experimental systems are vehicles for extension and enhancement. The SRS/E algorithm is a working and workable implementation of the Dynamic Expectancy Theory, but there is scope for additional capability. This section describes a small number of the possibilities.

7.1. An Association List

A component part of MacCorquodale and Meehl's interpretation of Tolman's *expectancy theory* proposed a separate sign to sign associative effect (denoted "$S_2S^\star$"). Such pairings may in particular record the association of arbitrary signs (S_2) to signs ($S^\star$) specifically identified as relating to desirable goal situations; the *secondary cathexis* postulate. The creation of a separate *Association List*, $\mathcal{A}$, within SRS/E would allow the attachment of multiple (secondary) goal states to a single (primary) goal definition. Signs detected as occurring concurrently with, or slightly preceding (giving a predictive element to the association) a predefined goal sign would be paired with the desired sign and this association saved on $\mathcal{A}$, figure 7-1. The strength of this association being subject to strengthening by *mnemonization* and weakening by extinction processes based on the frequency and temporal adjacency of the pairing.



Graphic 6.1 from monolith\dpmex.cdr

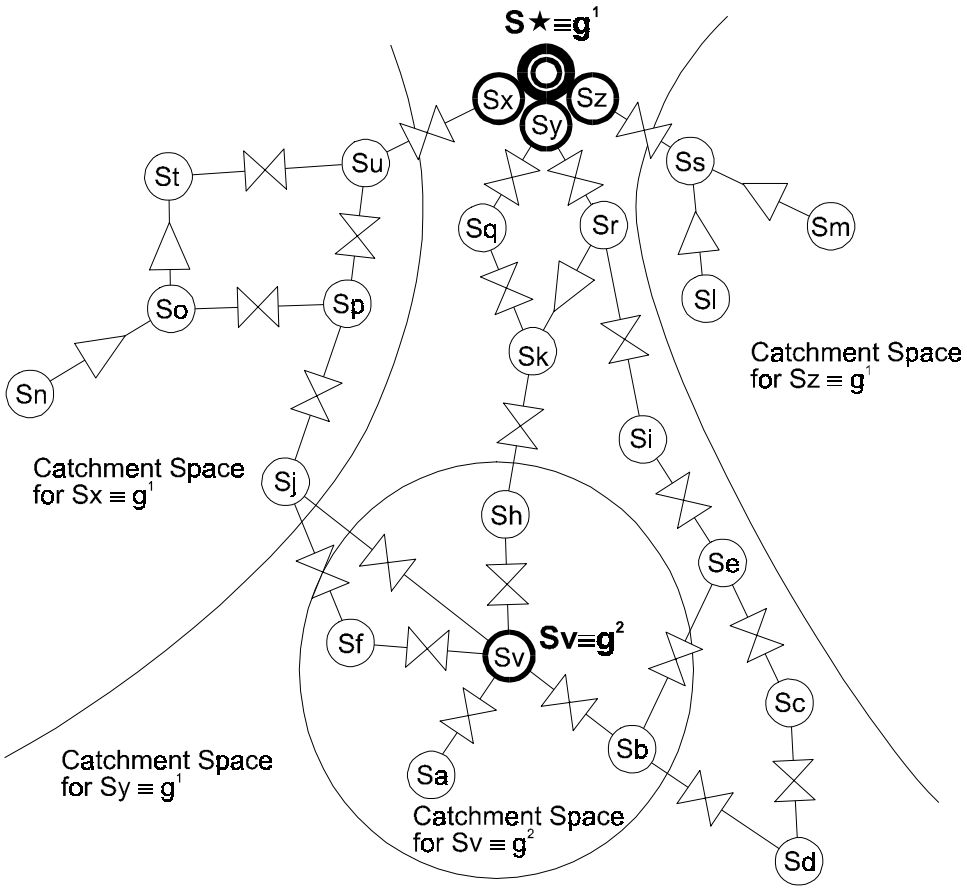
Figure 7-1: Sign-Sign Associations (Secondary Cathexis)

This arrangement allows greater flexibility in selecting goals from the Behaviour List, as there is no longer a requirement for the originator to identify specific tokens or signs to describe the goal. This form of association is different from the association phenomena described in the *classical conditioning* literature, in that it is not dependant on an unconditioned response (UR). The $S_2S^{\star}$ *sensory preconditioning* effect has been demonstrated under controlled experimental conditions, what relationship it may or may not have to classical conditioning phenomena is a matter of some conjecture. Bower and Hilgard (1981, pp. 330-331) review some of the evidence.

7.2. Seeking Multiple Goals Simultaneously

Multiple goals may be pursued in a more effective manner than the sequential strategy currently employed by SRS/E. Given several goals active on G, the SRS/E algorithm currently actively seeks the top-goal, and will pass secondary goals by, regardless of how close they are to the current path, or of the overall estimated cost of achieving the main goal and subsequently continuing to the secondary one. This was demonstrated in section 6.4. The algorithm normally takes the path of least estimated cost to the top-goal. Where a secondary goal is on the path, by either good-fortune or chance, then it is satisfied in passing.

Changes to the goal seeking process may be implemented either by building a single DPM where the action selected depends on both cost to goal and relative *goal priority*, or by computing several DPMs, and selecting an action on the basis of some, as yet undetermined *goal strength function*, *f*(`estimated_cost`, `goal_priority`), thus combining cost and priority. This would allow the animat to divert to secondary goals when they are close to the primary path. This, coupled to the proposed Association List, allows several paths to the desired specified goal state to be defined and pursued concurrently. Figure 7-2 illustrates the concept.



Graphic 6.2 from monolith\dpmex.cdr

Figure 7-2: Enhanced Goal Acquisition

In this example each goal (or goal by association) has a "catchment area", defined by the goal strength function. For each recomputation of the Dynamic Policy Map every sign in S will fall within the catchment area of one of the prioritised goals. So in this example if "Sb" was active ("Sb*"), the animat would use the μ-hypothesis "Hbv" to satisfy the lower priority goal g^2, even if the path "Sb*"-"Se" represented the lowest estimated cost path to g^1. The animat would then proceed to the

original g^1, possibly via the path "Sh"-"Sk" and so on. In the current implementation the animat selects from the available alternative paths "Sb*"-"Sv", "Sb*"-"Se" or "Sb*"-"Sd" entirely according to the lowest estimated cost path to g^1, and so may increase the total path to satisfy both goals unnecessarily. Even in the proposed regime the animat would pursue the path "Sd*"-"Sc" if that represents the lowest cost path, as "Sd" falls outside the catchment area defined for "Sv".

Goodwin and Simmons (1992) describe a decision theoretic approach to the balancing of multiple goals for a *HERO 2000* series mobile robot. Haigh and Veloso (1996) describe *Rogue*, a system for generating and executing plans with multiple interacting goals, where goal tasks may be interrupted or suspended.

7.3. An Explicit Template List

This extension to the SRS/E algorithm proposes an additional list type, the *Template List*, $\mathcal{T}$, to record the pattern of signs and actions used to build a new μ-hypothesis. Templates may at first be created at random, much in the manner that μ-hypotheses are in the present version of SRS/E. After a period of corroboration the effectiveness of each template may be assessed by reference to the confidence measures of the μ-hypotheses it was responsible for creating. Future bias being then given to those templates that are demonstrated to give rise to successful μ-hypotheses. This *meta-level learning* may be instrumental in explaining *learning-to-learn* phenomena described in the natural learning literature (although these phenomena may also be in part due to an increase in overall competence). The provision of a Template List would further allow the originator to bias the learning strategy of the animat according to pre-conceived notions of an intended environment or behavioural strategy.

The provision of a separate Template List equates, in some small measure, to *Popper's* notion of a "theory". Individual μ-hypotheses are generated from these meta-level objects, and in turn these meta-level objects may be judged according to the performance of their generated descendants.

7.4. Directing Learning Effort

The SRS/E algorithm is an implementation of an expectancy theory, reinforcement for individual μ-hypotheses is contingent upon their effectiveness as a predictive element. This reinforcement is not, in the system and experiments so far described, contingent on any notion of the value (as defined in the ethogram or elsewhere) in achieving goals defined for the system. There is a huge body of evidence that learning is indeed contingent upon the achieving a "desired" outcome (i.e. one which "reinforces".) An absolute distinction between predictive outcome and desirability is therefore an unnecessary one, and ultimately potentially disadvantageous to the system.

MacCorquodale and Meehl (1953, pp. 238-239) suggest increasing the expectancy-growth strength to a greater rate according to valence level. This is equivalent to increasing the value of the learning rate parameter α when a reward is detected as a result of satisfying a highly valenced prediction. In practice, adopting this strategy will have only a marginal effect on the system's overall observable behaviour. It also serves to confound two quite separate issues - the reliability of an expectancy and the usefulness of an expectancy. The reliability (as reflected in the various confidence measures) of the μ-hypothesis is properly determined by the ratio of successful to unsuccessful predictions, as has been the case. If an outcome is useful, then emphasis should be placed on the acquisition of μ-hypotheses that achieve it either directly or indirectly.

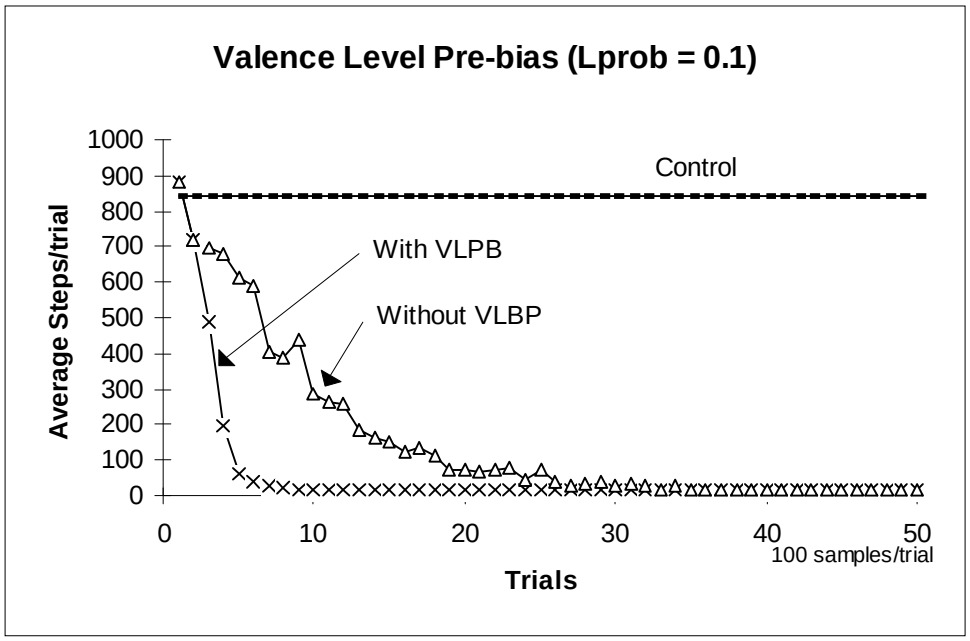
Each sign in $\mathcal{S}$ may therefore be graded according to the highest valence level it has achieved in the past in various Dynamic Policy Maps created by the system. Therefore, if a sign $\mathscr{s}$ has been nominated as a goal in the past, the learning sub-system should always create a new μ-hypothesis if the opportunity arises. If the sign $\mathscr{s}$ has been implicated at valence level two, then the learning system should be strongly biased to create a new μ-hypothesis, and so on, reducing as the highest recorded valence level for $\mathscr{s}$ falls away. In a practical system the probability of learning would reasonably be a function of (1) the highest ("best") valence level achieved by the sign; (2) the priority of the goal giving rise to the valencing; and (3) how recently the goal was valenced. Thus:

P(creation) ← *f*(`best_valence_level*goal_priority*recency_of_goal`)

eqn. (7-1)

Giving a situation where higher valence levels and greater goal priorities increase the probability that the unexpected occurrence of $\mathcal{S}$ will give rise to the formulation of a new μ-hypothesis to predict that sign. The probability further decreasing as time elapses since the goal was last asserted.

The current implementation of the SRS/E algorithm records the most significant valence level assigned to every element of the Sign List in the value `best_valence_level`. In an optional process to be referred to as *valence level pre-bias* μ-hypothesis creation by unexpected event (SRS/E step 8.2) always creates a new expectancy if the unpredicted sign has any valence level defined for it. This has no effect when the *learning probability rate* (Lprob) is 1.0, as all opportunities to learn are exploited unconditionally. The results of the experiments described in section 6.2 show the deterioration in learning performance as Lprob is reduced. Figure 7-3 compares the effect of enabling the valence level pre-bias option for the data in figure 6-2 (where Lprob = 0.1, Adisp = 1.0 and Arep = 0.0) against the original results.



monolith\results\vlpb\vlpb.xls

Figure 7-3: The Effect of Valence Level Pre-Bias

The dramatic improvement in learning performance is explained by the rate at which the valence level may propagate from the goal sign. With Lprob = 0.1, there is effectively only 10% chance that the crucial μ-hypothesis that connects the goal to a sign at valence level one will be created. Without this critical link, no DPM can be built, and goal seeking performance is restricted to random walk search. Once this link is created the catchment area within the DPM is widened and the corresponding random search time reduced.

The step-like performance shifts for many individual trials (which appear as the classical negatively decelerating learning curves when averaged over many trials) are a consequence of the abrupt connection of the growing network of latently learned expectancies, with those connected to the goal. By ensuring that the final connection is made (by pre-biasing it), and that the second connection is made on the next attempt, and so on, the portion of the graph connected to the goal is guaranteed to expand by at least one valence level on each trial. In figure 7-3 this would be a maximum of 14 trials. In practice this is reduced to around half this figure due to latent learning of the graph made during the trial-and-error search period of each trial.

7.5. Aversion

The discussion of SRS/E up to this point has only considered goals that are actively sought, and has not included situations where an action is to be avoided as it may lead to an undesirable outcome. There is a considerable body of evidence (Campbell and Masterson, 1969; Schwartz, 1989, Ch. 6) that animals and humans will actively avoid situations leading to certain sensations, variously described as undesirable, unpleasant or *painful*. The mechanism by which sensations are characterised in these ways in nature is not entirely clear.

For the purposes of the SRS/E algorithm it is sufficient to designate certain sensations, as encoded as input tokens or signs, as undesirable. This is a function of the ethogram design. μ-Hypotheses that predict the occurrence of these outcomes may be disadvantaged by additional cost estimates. The degree of this disadvantage being related to the given degree of undesirability of the resulting sensation, and

the confidence with which the outcome is predicted. It may be inappropriate to chain these aversions, in the manner of the positive goal seeking activities, as this may lead to a form (or analogue) of a *phobia*. Actions are avoided on the basis they might lead to an undesirable outcome at some time in the future, irrationally, as many actions may be taken to easily avoid the undesirable outcome. Clinical symptoms of phobias in humans seem unlikely to be related to this mechanism.