# Goal-Proximity Decision Making: Who needs reward anyway?

**Vladislav D. Veksler** (vekslv@rpi.edu)
**Wayne D. Gray** (grayw@rpi.edu)
**Michael J. Schoelles** (schoem@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180 USA

**Abstract**

Reinforcement learning (RL) models of decision-making cannot account for human decisions in the absence of prior reward or punishment. We propose a mechanism for choosing among available options based on goal-option association strengths, where association strengths between objects represent object proximity. The proposed mechanism, Goal-Proximity Decision-making (GPD), is implemented within the ACT-R cognitive framework. A one-choice navigation experiment is presented. GPD captures human performance in the early trials of the experiment, where RL cannot.

**Keywords:** RL, GPD, reinforcement learning, associative learning, latent learning, ACT-R, information scent, decision-making, seeking behavior, navigation, model-tracing.

## Introduction

How does a cognitive agent choose a path of actions from an infinitely large decision-space? Reinforcement learning (RL) models, which are models of human trial-and-error behavior, explain how an agent may reduce its decision-space over time by attending to the reward structure of the task-environment. However, as goals change, so does the reward structure of the agent's world. Relearning the reward structure for every possible goal may take an extremely long time. For greater efficiency, a cognitive agent should be able to learn more about its environment than just the reward structure, and to exploit this knowledge for achieving new goals in the absence of prior reward/punishment. For example, a person may see a hardware store on their way to the mall, and incidentally learn its location. Some time later, if they need to go to a hardware store, the person can find their way to that store, because they know its location. There had been no reward or punishment for the actions leading to this hardware store, and so the ability to find its location cannot be explained solely through the principles of reinforcement learning.

We propose a mechanism for making decisions in the absence of prior reward or punishment, and provide initial tests of its fidelity and efficiency as compared to RL. Given multiple possible paths of action, the proposed mechanism chooses the path most strongly associated with the current goal, regardless of prior reward. Strength of association between any two items, in turn, depends on experienced temporal proximity of those items. From here forth we refer to the proposed mechanism as GPD (goal-proximity decision-making).

The rest of this paper describes a key theoretical problem for RL models of decision-making (the 2-goal problem), briefly summarizes classic evidence in psychological literature for reward-independent decision-making in humans and animals, and presents two computational models that exemplify non-RL-based decision-making. We then outline the implementation of the GPD mechanism within the ACT-R cognitive framework. Finally, we describe a single-choice navigation experiment, and provide fits of GPD and RL decision mechanisms to human data. We conclude that GPD can account for human performance where RL cannot – prior to any reward or punishment.

### What this paper is not about

Because everything in cognition is so closely knit, the GPD theory may evoke topics that are outside of the scope of current work. The following topics are important to cognitive science but tangential to the focus of this paper.

First, GPD is not meant to replace RL, but rather to complement it. How GPD and RL may interact is a topic for further research.

Second, GPD does not address planning. GPD is a theory of immediate behavior; how this behavior may be used in complex planning procedures is a tangential topic.

Third, GPD partially addresses episodic memory and associative learning. However, associative learning is not the focus of this paper. Rather, the focus here is on the goal-oriented decision-making that can emerge from a simple associative learning mechanism. The topic of associative learning should comprise other lines of research (e.g. sequence recall, free association, priming) in addition to this one, and is too extensive to address here.

Fourth, GPD describes how an agent may choose which option to approach given multiple possible paths. Although avoidance behavior is just as important as approach behavior, and should eventually become part of the GPD theory, it is assumed here to be a separate topic.

### The 2-goal Problem

Consider a scenario where an agent has to achieve goal A, then goal B, in the same environment. To increase efficiency humans and animals would learn the environment during task A, and perform faster on task B (the Experiment below provides evidence for this phenomenon). That is, we do not just learn the positive utility for the actions that

helped us reach the goal, or the negative utility for the actions that failed to reach the goal; we also pick up on other regularities in the environment that may help us with possible future goals. RL-based architectures will have a problem matching human performance on this 2-goal problem.

To make this example more concrete, imagine how an RL-based agent may perform on a specific 2-goal problem. In this example, the first goal, A, can be accomplished by executing actions 1, 2, and 3. After trying the following sequences of actions, 1-2-4, 1-5-7, 1-4-3, finally the sequence 1-2-3 is attempted. Upon reaching the desired goal A, actions 1, 2, and 3 will be positively reinforced. The utility value of actions 1, 2, and 3 will increase every time that A is reached via this route, and soon these actions will fire without fail, greatly improving the agent's time to reach the goal.

Now imagine the task switches so that the agent has to find B in the same task environment. The shortest path to B would be to fire actions 1, 5, and then 7. Although the agent had previously reached state B, actions leading to this state were not positively reinforced because B was not the goal at the time. Thus, when presented with this new goal, RL performance will be at chance level.

RL, by definition, learns only the reward structure of the world, ignoring the rest of the environmental contingencies (with the exception discussed in the Model-based RL section below). In those cases where this ignored information may help in achieving new goals, it would be useful to have an additional mechanism for collecting and using this information (especially in the case of humans, where memory is relatively cheap as compared to additional trials). The mechanism proposed in this paper, GPD, should serve as such a complement for RL-based architectures.

## Background

Stevenson (1954) provided evidence that children are capable of resolving the 2-goal problem. In this study children were placed at the apex of a V-shaped maze, and the goal items were located at the ends of the arms of the V. Children were asked to find some goal-item A (a bird, flower, or animal sticker), and later asked to find a new goal B (a purse or a box). Although children were never rewarded for finding B, and did not know that they would be asked to look for it at any point, once presented with this goal, they proceeded to the correct arm of the maze more than 50% of the time.

This paradigm, called latent learning, does not just provide evidence that learning occurs in the absence of reward/punishment, but also that, given a goal, the learned information is reflected in decision-making, and ultimately in performance. Tolman provided evidence for latent learning in rats in the context of maze running (Tolman, 1948; Tolman & Honzik, 1930), and Quartermain & Scott (1960) displayed latent learning in human adults, substituting the maze environment for a cluttered cubicle shelf.

The following subsections describe Model-based RL – a RL framework that learns environmental contingencies beyond reward, Voicu & Schmajuk model of navigation – a model capable of resolving the 2-goal problem, and SNIF-ACT – a model that implements a decision mechanism similar to Voicu & Schmajuk within a unified cognitive framework.

### Model-based RL

Model-based RL (Sutton & Barto, 1998) extends RL by learning the environmental structure beyond action utilities. The term "Model" in "Model-based RL" refers to agent's internal model of the environment. An agent based on this framework is capable of planning its route before execution. However, the planning process itself is still based on RL. Using the example from the 2-goal Problem section, presented with a new goal B, and *having* the knowledge that 1-5-7 leads to B, a model-based RL agent will begin to plan its route by considering *random* actions. In other words, because this framework uses a decision mechanism based on RL, having the additional knowledge about the world does not reduce decision cycles.

### Voicu & Schmajuk

Although models of space navigation can employ RL (e.g. Sun & Peterson, 1998), there is a class of decision mechanisms employed in many artificial navigation systems that do not use RL representation (for review see Trullier, Wiener, Berthoz, & Meyer, 1997). As Trullier et al. state, "Navigation would be more adaptive if the spatial representation were goal-independent" (p. 489).

In a primary example of goal-independent representation Voicu and Schmajuk (2002) implemented a computational model that learns the structure of the environment as a network of adjacent cells. Once a goal is introduced, reward signal spreads from the goal-cell through this network, such that the cells farther from the goal-cell receive less activation than those that are close. Goal-driven behavior in this model comprises moving towards the cells with the highest activation.

Once this model memorizes the map of the environment, it does not need to learn the reward structure through trial-and-error; rather, the utility of each action-path is identified through spreading activation from the goal. In this manner, this model resolves the 2-goal problem.

One major limitation of this model is that it makes unrealistic assumptions about the world (e.g. that it can be neatly mapped out as a grid of adjacent spaces). This model would be computationally infeasible for sufficiently large, dynamic, probabilistic environments. Additionally, this model is not integrated within a larger cognitive framework. As a standalone model of maze navigation behavior in an oversimplified environment, there are questions as to the scalability and fidelity of the model. The following sections address how a similar mechanism, where decisions are based on spreading activation from the goal, may be

implemented within a unified cognitive framework, such that integration is at the core of modeling.

### SNIF-ACT

SNIF-ACT (Fu & Pirolli, 2007) is a model of human information-seeking behavior on the World Wide Web. The pertinence of SNIF-ACT to current work is that it is a model of how humans use declarative knowledge (rather than action utilities) in goal-driven behavior in a very rich and unpredictable task-environment. The World Wide Web is unpredictable in the sense that there is no way for any of its users to know what links they will encounter during web browsing. For this reason an agent must be able to evaluate its actions (which link to click) without any prior reinforcement of those actions.

The action of clicking a link in SNIF-ACT is based not on the previous reinforcement of clicking on that link, but rather on the semantic association of the text in the link to user goals (information scent). To implement this concept in ACT-R, Fu & Pirolli changed the utilities for clicking links based on the link-goal association strengths (note the similarity to the Voicu & Shmajuk model). This is different from the standard ACT-R implementation, where the decision mechanism is based on RL. Changing the utility mechanism in this way allows SNIF-ACT to make non-random decisions between multiple matching actions that have never been reinforced.

Besides being limited to text-link browsing, SNIF-ACT's other major limitation is that it does not learn the association strengths between links and goals, but rather imports these values from an external source. However, SNIF-ACT's decision-making mechanism is an excellent example of how to achieve goal-driven behavior in the absence of prior reinforcement within the ACT-R framework.

## Goal-Proximity Decision Making

RL cannot account for human/animal decision-making in the absence of reward. The Voicu & Schmajuk and the Fu & Pirolli models described above suggest an alternative decision mechanism where agent choice depends on spreading activation from the goal.

More specifically, these models employ reward-independent associative knowledge to represent environmental contingencies. The decision mechanism in both models works by approaching the option most strongly associated with the goal element.

In the Voicu & Schmajuk model, the strength of association between two elements is inversely proportional to the physical distance of those elements in space. In SNIF-ACT, the strengths of associations are imported from an external source – Pointwise Mutual Information engine (Turney, 2001), where association strength between two words is incremented every time that the two words co-occur within a window of text, and decremented every time that the two words occur in the absence of one another.

In other words, the experienced temporospatial proximity between items X and B may be employed to predict whether X is en route to B. While the agent is seeking some goal, A, it may be learning the proximity of elements in its environment, including the proximity of X and B. Given a new goal, B, the agent can use its knowledge to judge the utility of approaching X to find B. In this manner, the environmental contingencies learned while performing goal A can help to improve agent performance on goal B, thus resolving the 2-goal problem.

We call this mechanism Goal-Proximity Decision-making (GPD). In more generic terms, GPD (1) relies on having associative memory, where association strengths between memory elements represent experienced temporal proximity of these elements, and (2) chooses to approach the environmental cue that is most closely associated with its current goal.

### Implementation

We implement GPD in the ACT-R cognitive architecture (Anderson & Lebiere, 1998). ACT-R comprises a production system as the central executive module, a declarative memory module, a goal module, and visual and motor modules.

To implement GPD in ACT-R, we developed an ACT-R model that, given some goal G, looks through all the options on screen, performing retrievals from memory. Retrievals from memory in ACT-R, among other factors, depend on spreading activation from the goal – such that the memory elements that are more strongly associated with G are more likely to be retrieved. The GPD model then clicks on the last option to have been retrieved from memory.

Although ACT-R employs the spreading activation mechanism, making for an easy implementation of the GPD model (only 13 productions), it does not make predictions about how association strengths between memory elements are learned. ACT-R 4.0 (an older version) had a mechanism for associative learning (Lebiere & Wallach, 2001; Wallach & Lebiere, 2003). However, according to Anderson (Anderson, 2001), this particular form of associative learning turned out to be "disastrous", and produced "all sorts of unwanted side effects" (p. 6).

To implement associative learning in ACT-R we first create an episodic buffer – a simple list containing the names of recently attended memory elements. Whenever the model checks the contents of the visual buffer (visual attention), the name of the memory element from the visual buffer is pushed into the episodic buffer.

Next, we update association strengths between the latest episode and every other item in the episodic buffer. To do this we employ error-driven learning. Error-driven learning, also known as the Delta rule, is widely accepted as a psychologically and biologically valid mechanism of associative learning (for psychological, computational, and biological review of error-driven learning see Gluck & Bower, 1988; O'Reilly & Munakata, 2000; Shanks, 1994). For each new element $j$ and previously experienced element $i$, the strength of association between $j$ and $i$, $S_{ji}$, at current time, $n$, is increased in the following manner:

$$\Delta S_{ji}(n)=\beta[a_i(n)-S_{ji}(n-1)]$$

where $\beta$ is the learning rate parameter, and $a_i$ is the activation of each element $i$ in the episodic buffer. Episodic activation, $a_i$, is assumed to decrease by some decay parameter, ɔ, at each tic. It should be noted that we did not employ the ACT-R native constraints for memory activation and decay – ACT-R memory decay implementation accounts for frequency, recency, and spreading activation, bearing peripheral complexity, to be examined at a future date. The pseudocode for the GPD model and this associative learning mechanism is provided in Table 1.

Table 1. Implementation of GPD.

Sji is the association strength between memory elements j and i
ɔ is the rate of decay of activation of objects in episodic memory
β is the associative learning rate parameter

```
########################################
# GPD algorithm
given a goal, G, and current best option, Y {
  for each option in the environment, X {
    learn episode (X)
    given two options, X and Y {
      attempt retrieval from declarative memory
      spreading activation from G
      set Y to be the retrieved memory element
    }
  }
  learn episode (Y)
  approach option Y
}

########################################
# Episodic/associative learning
learn episode (j) {
  activationOfItem = ɔ
  for each item in episodic-buffer, i {
    Sji += β * (activationOfItem - Sji)
    activationOfItem = activationOfItem * ɔ
  }
  push j into episodic-buffer
}
```

## Experiment

The purpose of this experiment is to collect data for validation of how GPD can account for human choice where RL cannot. The structure of the experiment reflects the 2-goal problem. More precisely, this experiment requires the participants to traverse a simple maze in search of different goal-items presented one at a time. Whereas RL would predict that reward structure is updated after the agent reaches a goal or a dead-end, GPD would predict that the agent also learns where other items in the maze are located. When asked to find a *new* goal, RL should perform at chance level (since there has been no reward for this goal), whereas GPD should perform above chance level. Human data from this experiment should provide a stark contrast between the two decision mechanisms.

**Participants**

Twenty-one human participants, consisting of undergraduate students at RPI, were asked to participate for course extra credit, as specified by course instructor.

**Materials**

The experiment was presented as a point-and-click application on a 17" computer screen, set to 1280x1024 resolution. Participants were presented with 150x200 pixel option buttons, where each button displayed either a letter from the English alphabet, or one of the symbols shown in Figure 1.

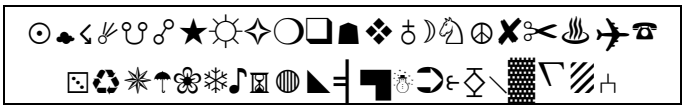

Figure 1. Stimuli used for 3-choice mazes.

**Procedure and Design**

The experiment employed a single-group design with no between-subject variables. Participants were asked to perform a simple exploratory maze navigation task. Each participant had to complete two 2-arm mazes (2 arms, 2 goal items in each arm) and four 3-arm mazes (3 arms, 3 goal items in each arm) in the following order: 2-arm, 3-arm, 3-arm, 2-arm, 3-arm, 3-arm. The choice and goal items in each of the 2-arm mazes were random letters of the English alphabet, and the choice and goal items of the 3-arm mazes were symbols randomly chosen from Figure 1. Participants were required to continue with a given maze until they completed 6 consecutive error-free trials (trials where only the correct path to the goal was taken) in the 2-arm mazes, or 12 consecutive error-free trials in the 3-arm mazes.

For each trial, participants were asked to find one of the goal items (for example, in the maze displayed on left of Figure 2, a goal could be: C, D, E, or F), such that no two successive trials would have repeating goals. The idea here is to replicate the 2-goal (or rather *n*-goal) problem design – while participants are looking for a given goal item they may be learning the maze, and will be able to perform above chance-level when presented with the next goal item.

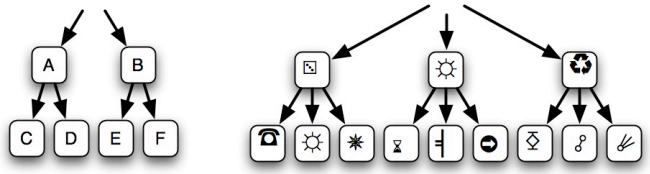


Figure 2. Sample navigation mazes, 2-arm condition (left) and 3-arm condition (right).

*Trial Design:*

Each trial persisted until the participant found and clicked the required goal item. At the beginning of each trial, participants were presented with the top-level options. After choosing one of top-level options, participants were presented with the bottom-level options (for example, in the 2-arm maze in Figure 2, a participant is first presented with options A and B, and if they choose option A, they are presented with options C and D). If the participant chose the wrong path to the goal, upon choosing one of the bottom-level options, they were presented with a "Dead End" screen, and taken back to the top-level options. If the

participant found and clicked their current goal item, they were presented with their next goal.

*Screen Design:*

To ensure that participants attended each option, the options were always covered with a grey screen until clicked. Another click was necessary to cover an uncovered option before proceeding. After the first option is uncovered and covered, a participant may proceed to uncover the next option. Once all options on screen have been viewed and covered, the participant could make their choice with an additional click. Additionally, participants were not be able to rely on their location memory, as the location of each option on screen was randomized; thus participants were forced to attend every item (i.e. the participant could not say, "when I go left, I get C and D," they had to recall that, "B leads to C and D," instead).

## Modeling

Human data were analyzed in terms of agreement with four models: GPD, RL, Random, and IdealPerformer. The Random model selected which option to click at random, and the IdealPerformer model remembered everything perfectly (which choices followed which other choices) and made choices with perfect memory. The RL model simply increased the utility of a goal-choice pair if the choice led to the goal successfully, and decreased it otherwise; the option with the highest utility warranted a click (no noise was added), and if multiple options had the same utility, the choice was random. After a few (less than 10) variations were attempted, the best-fit GPD model was derived to have error-driven learning with the following parameters: ɔ=.5, $\beta$=.01. No noise was added to spreading activation.

Model data was collected using the model-tracing technique (Anderson, Corbett, Koedinger, & Pelletier, 1995, as cited by Fu & Pirolli, 2007). For each human participant, for each decision, each model was provided with the same experience as the human participant up to that choice point, and then model's would-be choice was recorded. For example, imagine that Table 2 presents data for a human participant having gone through the maze shown on left of Figure 2. At the bolded choice-point (trial 1), being that there is no experience with the maze, all models would choose randomly. Let us say that both the RL and the GPD models chose B. Thus, what will be recorded is that these two models made an error on trial 1, whereas the human participant did not. However, the experience added to the two models will be based on human choice. At the end of trial 1, RL will have learned that the D-A (if goal is D, click A) goal-choice pair has a positive utility. GPD will have learned that D is strongly associated with C, less so with A, and even less with B, and that C is strongly associated with A, and less so with B. At the underlined choice point (trial 2, top), the RL model will still have to make a random choice (utilities for C-A and C-B goal-choice pairs are both 0 at that point). The GPD model, having learned that C is more associated with A than with B, will choose A.

Table 2. Sample data log for a human participant.

```
Trial 1: goal=D:
  looked at A, looked at B, clicked A,
  looked at C, looked at D, clicked D, success
Trial 2: goal=C:
  looked at B, looked at A, clicked B,
  looked at E, looked at F, clicked F, fail
  looked at B, looked at A, clicked A,
  looked at C, looked at D, clicked C, success
…
```

## Results and Simulation

Each model's performance was averaged over 10 model runs for each decision point. Results from the first 2-arm maze were ignored as training data. Results for human and model performances on the first choice of each of the first 6 trials for the other 2-arm maze (maze 4) are shown at the top of Figure 3 (only the first 6 trials are shown because some participants did not have data beyond the 6th trial). Results for human and model performances on the first choice of each of the first 14 trials for the 3-arm mazes (averaged over all mazes: mazes 2, 3, 5, and 6) are shown at the bottom of Figure 3 (only the first 14 trials are shown because some participants did not have data beyond the 14th trial).

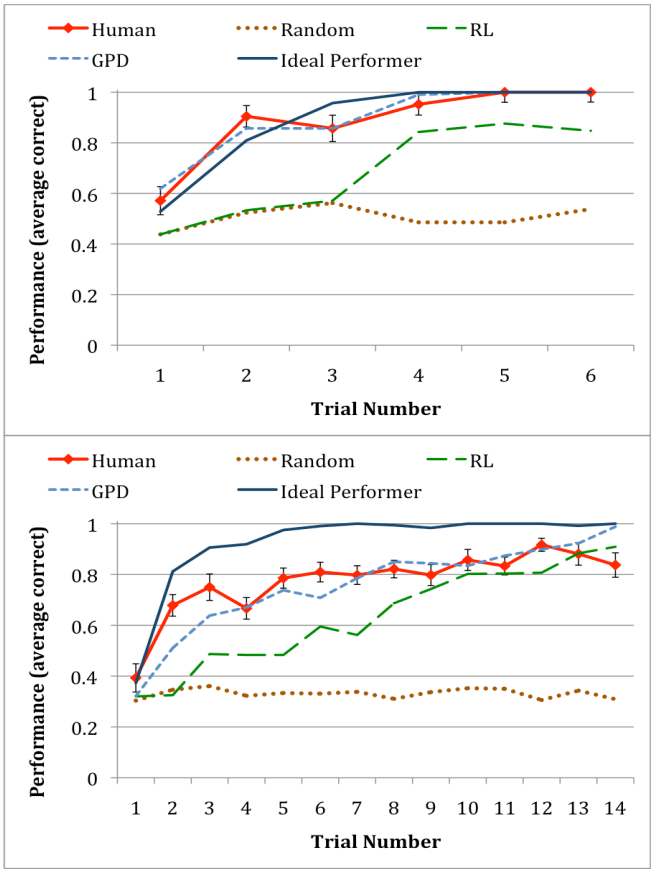


Figure 3. Average performance from human participants, GPD, RL, Random, and IdealPerformer models on the 2-arm maze (top), and the 3-arm mazes (bottom). Error bars represent standard error based on 21 participants.

Table 3. Root mean square error (RMSE) between human and model performances, by trial.

| | 2-arm | 3-arm |
|---|---|---|
| GPD | 2.07% | 7.95% |
| RL | 14.84% | 18.29% |
| IdealPerformer | 4.07% | 16.34% |
| Random | 45.32% | 45.79% |

Table 3 displays Root Mean Square Errors (RMSE) between average human and model performances for the data displayed in Figure 3 – performance on the first choice of each trial for the first 6 trials of the second 2-arm maze, and the first 14 trials of the four 3-arm mazes.

The key aspect to focus on is the early part of the curves in Figure 3, where RL simply cannot account for human-level performance. IdealPerformer model assumes that associations between the clicked top-level choices and their respective bottom-level objects are strengthened, and that the non-clicked top-level choices do not interfere. For example, on trial 1 shown in Table 2, the IdealPerformer model will have only learned the association between the clicked option, A, and the ensuing options, C and D. GPD, however would increment association strengths between C/D and all of their preceding items: both A and B. Thus, IdealPerformer learns unrealistically fast, and RL learns unrealistically slow.

## Summary

Whereas reinforcement learning accounts for human decision-making based on prior reward, this paper proposes a mechanism to account for human choice in the absence of reward, based on associative learning. The proposed mechanism, GPD, was implemented in the ACT-R cognitive architecture, and examined in its ability to simulate human behavior in a simple forced-choice navigation task. GPD was able to account for human data where RL could not – in the beginning of the task, before reward or punishment for finding a given goal could have been presented.

To implement GPD in the ACT-R cognitive architecture, it was necessary to add two things. First, we wrote an ACT-R model that made retrievals based on spreading activation from the goal, and clicked on the retrieved option. Second, associative learning was introduced: keeping recently attended memory elements in an episodic buffer, and using error-driven learning to increase the strengths of association between memory elements based on their proximity in the episodic buffer.

GPD seems to be a necessary supplement to RL for explaining human decision-making. We are currently in the progress of using GPD to play Tic-Tac-Toe, providing initial grounds for the claim that GPD can be used in more than just navigation tasks, but rather in navigating any decision-space, including board games. We are also beginning to explore how this mechanism scales to more complex, dynamic task environments (e.g. exploration of Second Life virtual worlds).

In addition to testing GPD with board games and exploration of virtual worlds, it will be necessary to integrate GPD with RL, for more complete approach/avoidance behavior. Future studies will focus on integration of GPD with other cognitive mechanisms, and testing the integrated framework across a wide range of tasks.

## References


Anderson, J. R. (2001). *Activation, Latency, and the Fan Effect.* Presented at the Eighth Annual ACT-R Workshop, Pittsburgh, PA.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A Cognitive Model of User Navigation on the World Wide Web. *Human Computer Interaction*.

Gluck, M. A., & Bower, G. H. (1988). From Conditioning to Category Learning - an Adaptive Network Model. *Journal of Experimental Psychology-General, 117*(3), 227-247.

Lebiere, C., & Wallach, D. (2001). Sequence Learning in the ACT-R Cognitive Architecture: Empirical Analysis of a Hybrid Model. In R. Sun & C. L. Giles (Eds.), *Sequence learning : paradigms, algorithms, and applications*. New York: Springer.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience : understanding the mind by simulating the brain*. Cambridge, Mass.: MIT Press.

Quartermain, D., & Scott, T. H. (1960). Incidental learning in a simple task. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 14*(3), 175-182.

Shanks, D. R. (1994). Human associative learning. In N. J. Mackintosh (Ed.), *Animal learning and cognition.* (pp. 335-374). San Diego, CA: Academic Press.

Stevenson, H. W. (1954). Latent Learning in Children. *Journal of Experimental Psychology, 47*(1), 17-21.

Sun, R., & Peterson, T. (1998). Autonomous learning of sequential tasks: Experiments and analyses. *IEEE Transactions on Neural Networks, 9*(6), 1217-1234.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts: The MIT Press.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology, 4*, 257-275.

Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J. A. (1997). Biologically based artificial navigation systems: review and prospects. *Prog Neurobiol, 51*(5), 483-544.

Turney, P. (2001). *Mining the Web for synonyms: PMI-IR versus LSA on TOEFL.* Presented at the Twelfth European Conference on Machine Learning, Berlin: Springer-Verlag.

Voicu, H., & Schmajuk, N. (2002). Latent learning, shortcuts and detours: a computational model. *Behavioural Processes, 59*(2), 67-86.

Wallach, D., & Lebiere, C. (2003). Implicit and explicit learning in a unified architecture of cognition. In L. Jimenez (Ed.), *Attention and implicit learning.* (pp. 215-250). Amsterdam, Netherlands: John Benjamins Publishing Company.