Using DSHM to Model Paper, Rock, Scissors

Matthew F. Rutledge-Taylor (mattrt@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Robert L. West (robert_west@carleton.ca)
Institute of Cognitive Science, Department of Psychology, Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, K1S 5B6, Canada

Abstract

Dynamically Structured Holographic Memory (DSHM) is architecture for modeling memory. It was originally designed to account for long-term memory alone. However, the current model of Paper, Rock, Scissors (PRS) play (and Rock=2 PRS play) provides evidence that DSHM can be used to model tasks where the truth of facts change quickly and dynamically.

Keywords: DSHM, memory, game playing, decision making.

Dynamically Structured Holographic Memory

Dynamically Structured Holographic Memory (DSHM) is an architecture for modeling memory. It was designed to account for how stable information is stored and retrieved from long-term memory. DSHM does not presume that the relationships between concepts are entirely static. Rather, it explicitly accounts for how concepts can evolve to form associations of different strengths, over time. DSHM has been used successfully to model the fan effect (Rutledge-Taylor & West, 2008; Rutledge-Taylor, Pyke, West and Lang, 2010). It has also been used as the basis for a recommender system (Rutledge-Taylor, Vellino & West, 2008). In both of these applications, DSHM commits a set of static facts to memory.

DSHM was not designed as a store for information whose relevance is short-lived and potentially contradictory to new information. This is because there are no native mechanisms that cause old memories to decay, or otherwise contribute less than recent memories in decision making. This sort of information, whose relevance is time dependent, is not uncommon in strategic decision making tasks, such as those in that take place in competitive games described by game theory (VonNeumann & Morgenstern, 1944). For example, in the game Paper, Rock, Scissors (PRS), successful players are able to detect and exploit repeating patterns of sequences of moves in the play of opponents (West & Lebiere, 2001). Additionally, they must adapt to ignoring old biases in opponents play and discover new biases.

The fact that DSHM was not designed to model these sorts of dynamic memory tasks does not preclude the possibility that it might be used to build successful models of human performance in these sorts of games. This possibility was examined by building DSHM models of human performance in the game PRS and a modified version of PRS. These models are presented herein.

Existing Models of PRS

In order to provide some base-level of expectation for what might constitute good performance in models of PRS play, some existing models are briefly reviewed.

Perceptron Models

It has been shown that simple perceptron-like neural networks can be used to model human behaviour in standard PRS games (West, 1998; West & Lebiere, 2001). The networks take sets of past opponent moves as input, and provide choices of next move as output. The networks are constructed such that they each have an output layer of three nodes, one corresponding to each of the three play options: paper, rock, and scissors. Each has one or more groups of input nodes. Each group includes three input nodes, one for each play option. Each input node is connected to each output node. The connections between nodes are assigned integer values (or, weights), which start at 0.

If a network has only a single input group, it takes only its opponent's last move as input. To determine what option to play, the connections between the node in the input group corresponding to the opponent's move and each of the output nodes are compared. The output node attached to the connection with the greatest value determines which move the network selects (ties are decided randomly). If the network's decision results in a win, the relevant connection is rewarded by increasing its value by one. If the result is a loss, the connection is punished by reducing its value by one. Ties are treated differently by two variations of this basic network design (West & Lebiere, 2001). Networks called 'passive' treat ties as neutral events and neither reward or punish the connection values after a tie. Networks called 'aggressive' punish connections leading to ties by 1.

Networks with two or more input groups take a set of the opponent's last moves as input. Each additional input group beyond the first corresponds to a move further back in the opponent's play history. For example, with two input groups, one corresponds to the opponent's last move as input, while the other takes the opponent's second to last move as input. For these networks the output is determined by summing the connections between one node from each input group (corresponding to the move played on that past occasion) and each output node. Rewards and punishments are applied to all connections that contribute to the output decision. In addition to being labelled as either passive or

aggressive, these networks were also labelled as 'lag 1', 'lag 2' or 'lag 3' depending on whether they attended to opponents' last one, two, or three past moves.

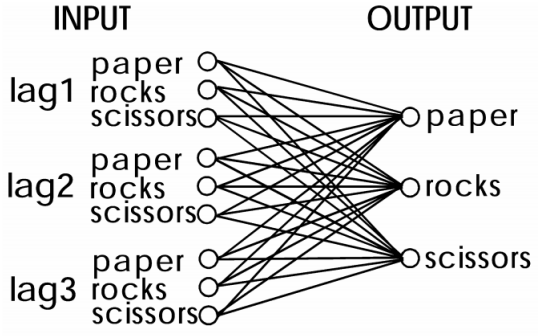

Figure 1: Perceptron Networks

Both West (1998), and West and Lebiere (2001) concluded that the aggressive lag 2 network provided the best model of human PRS play. Both the humans and the aggressive lag 2 networks were able to beat the passive lag 2 and aggressive lag 1 networks (West, 1998; West & Lebiere, 2001), by statistically significant margins. On average, humans lost to the aggressive lag 2 networks by a small margin (West & Lebiere, 2001). However, the authors suggest that this may be due to imperfect attention and motivation on the parts of the human participants.

Perceptron Rock=2 The standard PRS game played by humans and network models in West (1998) and West and Lebiere (2001) were perfectly symmetrical. The three play options were identical in that each beat one of the other two moves and lost to the other; a rock versus scissors win was no different from a scissors versus paper win. In Rutledge-Taylor and West (2004), a modified version of PRS, where rock versus scissors wins were worth two points, while the other two outcomes were worth only one point each, was investigated. Ten human participants played one game against each of three network opponents. The network opponents were: the aggressive lag 1, the aggressive lag 2, and a 'rock=2' lag 1. The rock=2 network rewarded network connections by two when it won with rock. Table 1 presents the mean points differences in the final scores of games between the human participants and each of the network models. All games consisted of 300 trials each.

Table 1: Humans versus networks

Network Model	Mean Pts. Diff.
Agg. Lag 1	16.5
Agg. Lag 2	5.7
Rock=2 lag 1	25.6

Two conclusions were drawn from this experiment: 1) humans were able to take advantage of the fact that wins using rock were worth two, while all three network opponents were not; 2) the rock=2 network performed the worst of all the network models due to the fact that rewarding rock wins by two became a liability (and not the anticipated advantage). This larger reward unbalanced the reward system in such a way that the rock=2 network played rock too frequently, and this was exploited by the human players.

An additional observation was that in Rock=2 PRS the frequencies with which a player played each of the possible moves did not predict the game's final scores. For example, if two players each play paper, rock and scissors exactly 1/3 of the time each, they will tie, on average, if they are playing randomly. However, it has been demonstrated that human players (and the network models) do not play randomly (West, 1998; Rutledge-Taylor & West, 2004).

Rather than playing randomly, superior players exploit weaker opponents by predicting their opponents' moves and making winning moves. As a result, they are able to achieve higher win rates than would be predicted by play probabilities alone. This effect is particularly important in the Rock=2 game, as being able to orchestrate rock versus scissors plays and to be able to avoid the opposite play is crucial to success in this game. Thus, a measure called the strategy index was invented.

The strategy index is calculated according to formula 1, below. Given two players, player 1's strategy index is player 1's total points scored minus player 2's total points, divided by the number of games played. The predicted points difference is calculated using game theory (i.e., each player is assumed to have played randomly according to probabilities determined by the actual ratios with which the options were chosen). A positive strategy index indicates a superior ability to correctly anticipate opponent's plays and achieve a higher than probabilistically predicted number of points. The raw strategy index is relative to the number of trials per game. So, it can be also represented as a percentage according to formula 2.

(1) Strategy index = average actual points difference per game – predicted points difference per game
(2) Strategy index percentage = Strategy index / number of trials per game

For example, if two players play a game of 300 trials, and each player plays paper 100 times, rock 100 times, and scissors 100 times, game theory predicts that they will tie (on average). However, it is possible for one player to win all 300 trials by always matching the opponent's move with the move that beats it. In this case, the winning player would score a perfect strategy index of 300 (or 100%). In real games, a strategy index percentage of 3% or more is considered very good.

In Rutledge-Taylor & West (2004) a network model of human Rock=2 play was created by using a genetic algorithm to find a reward matrix that resulted in human like play. The criterion was to match as closely as possible the human mean points difference and the mean strategy indices

against the three opponents in table 1. The best reward matrix was the following: rock wins = 3, paper wins = 2 scissors wins = 0; rock tie = -1, paper tie = -1, scissors tie = 0; and -3 for all losses. The performance of this model is presented below.

ACT-R Models

ACT-R models of both standard PRS (Lebiere & West, 1999) and of Rock=2 PRS (Rutledge-Taylor & West, 2005) have been created. Both employ an exemplar based approach and manipulate noise to establish a best fit.

For both models a chunk type with four slots was used. The isa slot was tagged with PRS to indicate that is was a PRS relevant chunk. The three remaining chunks encoded a sequence of three moves, by the model's opponent: lag0, is the opponents current, or predicted move; lag1 is its previous move; and, lag2 is its second to last move. An example is illustrated in figure 2.

Goal
isa PRS
lag2 Paper
lag1 Rock
lag0 nil

Figure 2: Example chunk

When the model's opponent makes a move, a chunk encoding its last three moves is put into the goal, and then popped to make it a chunk in memory (either creating a new chunk or reinforcing an existing chunk). To predict the opponent's move, the model attempts to retrieve a chunk from memory that matches the opponent's last two moves (slots lag1 and lag2). The value of the lag0 slot is the move the model predicts its opponent to make. It then plays the move that beats it.

In Lebiere and West (1999), both humans and the lag 2 ACT-R model were pitted against lag 1 and lag 2 versions of the ACT-R model. Consistent with the findings in West (1998), both humans and the lag 2 ACT-R model were able to beat the lag 1 opponent; however, exact scores were not reported.

ACT-R Rock=2 Several ACT-R models of human Rock=2 PRS play were presented in Rutledge-Taylor and West (2004). These models were similar to those appearing in Lebiere and West (1999), however, they differed in that they were designed to be sensitive to the unequal payoffs in the Rock=2 game. This sensitivity was achieved by reinforcing certain kinds of chunks more than others depending on what the opponent's last play was. This was done by harvesting these chunks twice. Rutledge-Taylor & West (2005) tested three variations on this strategy. One model paid extra attention to cases when its opponent played rock; another attended more closely to scissors; while the third attended more closely to both rock and scissors (effectively paying less attention to paper).

The result was that the third model provided the best match to the human data. This makes intuitive sense in that a human player is likely to incorporate a defensive component to his or her game, which his to be wary of when the opponent is likely to play Rock; however, he or she might also incorporate an offensive component which is to also focus on when the opponent might play scissors. Winning with Paper, or losing to a Paper play, is a less important event in the game.

DSHM PRS Models

Given the broad similarities between DSHM and the declarative memory system of ACT-R (Rutledge-Taylor & West, 2008), the DSHM models here were based on the ACT-R models described above.

The DSHM models took sequences of opponent's plays, encoded as ordered complex items. The items consisted of two, three, or four atomic items, for lag 1, lag 2 and lag 3 models respectively; the extra item is the predicted or current play by the opponent. The right-most item represented the opponent's last move, while items to the left represented previous plays. For example, if the opponent's last few plays were:

…, rock, paper, paper, rock, scissors;

a lag 1 DSHM model would learn the following pattern after scissors was played:

[rock:scissors];

thus, reinforcing the association between 'scissors' as a play the follows 'rock'. A lag 3 DSHM model would learn the pattern:

[paper:paper:rock:scissors].

An interesting architectural point that should be made here is that DSHM reinforces all of the combinations of consecutive sets of sub-items in the input, when learning ordered complex items. Thus, given the lag 3 input above, the following sequences of items are reinforced after scissors is played:

[paper:paper:rock:scissors],
[paper:rock:scissors],
[paper:paper:rock],
[rock:scissors],
[paper:rock], and
[paper:paper].

So, in a sense, DSHM models of two or more lags incorporate some of the learning of shorter lagged models as well. An additional consideration is that this results in a potential liability for DSHM, as shorter sequences receive repeated reinforcement for several consecutive trials. In this example, this will have been the third time that

[paper:paper] had been reinforced: the third time immediately after the play of 'scissors'; the second time, when 'rock' was played; and the first time when the most recent of the two 'paper' plays was made. It is also the second time that [paper:paper:rock] will have been reinforced (the first time being when 'rock' was played). Thus, in this respect, DSHM models of PRS differ somewhat from both the perceptron-like networks, and the ACT-R models discussed above.

Rock=1 Simulations

A variety of DSHM PRS players were built. The manipulated parameters were: number of lags and the length of the vectors used to represent items. It would not be appropriate to embark on a complete discussion of the inner-workings of DSHM here (see Rutledge-Taylor, Pyke, West & Lang, 2010; Rutledge-Taylor & West, 2008). For now, it is sufficient to understand that vector length is correlated with memory capacity. Additionally, lower vector lengths contribute to noise-like effects due to an increased amount of interference in the system.

Each combination of two sets of values for these parameters was used to build a unique DSHM PRS player:

Lags: [1, 2, 3];

Vector lengths: [32, 64, 128, 256, 512, 1024, 2048].

Each DSHM model played against the aggressive lag 1, aggressive lag 2, and passive lag 2 network models.

Evaluation To determine which DSHM performed most like human players, data from West and Lebiere (2001) was used as a comparison: human players average 9.99 (s.d. 19.61) more wins than the aggressive lag 1 networks, after 300 trials; lost to the aggressive lag 2 models by an average margin of 8.89 (s.d. 19.74), after an imprecise number of trials; and, beat the passive lag 2 by 11.14 wins after 287 trials.

Given that understanding human play against the aggressive lag 2 networks is difficult due confounding factors discussed in West and Lebiere (2001), and the fact that an exact target win difference (after a fixed number of trials) is not available, comparison to the data against this opponent was simplified.

The DSHM models were rated according to the mean squared difference between their average final scores against the aggressive lag 1 and passive lag 2 networks, and the average finals scores of humans against these models. Additionally, DSHM models that won against the aggressive lag 2 were disqualified as potential models of human play. This is because all that is certain about human performance against the aggressive lag 2 networks is that humans lost to these networks, on average. Additionally, West and Lebiere (2001) discuss factors that could make the interpretation of human players' performance against the aggressive lag 2 networks difficult (e.g, it is less fun to play and lose against a stronger opponent).

Results Of all the models tested, one produced results that came very close to the human data. The Lag 3 DSHM model with vector lengths of 1024 scored an average of 10.89 wins more than the aggressive lag 1 network, 13.24 more wins than the passive lag 2, and lost to the aggressive lag 2 by an average of 6.22 wins per game.

The fact that the best DSHM model was a lag 3, not a lag 2 model, was surprising at first. Lebiere and West (1999), and West and Lebiere (2001), found that lag 2 ACT-R and lag 2 network models provided the best fit to the human data. However, the fact that the DSHM lag 3 models incorporated lag 2 and lag 1 memory behaviour makes this result more consistent with previous findings. The DSHM lag 3 model weighs lag 1 sequences the most, lag 2 sequences second, and lag 3 sequences the least. So, it could be argued that, on average, the lag 3 DSHM models are more like lag 2 ACT-R and network models, than the lag 3 ACT-R and network models.

Rock=1 Model Comparison

This paper reviews the three different types of models of PRS play: ACT-R, DSHM, and perceptron-like networks. In each case, the model of human play, played games consisting of 300 trials against the aggressive lag 1 network, and games of 287 trials against the passive lag 2 network. For each model, the mean difference in the number of wins scored by the model and the opponent network was recorded. Humans scored, on average, 9.99 more wins than the aggressive lag 1 networks, and 11.14 more wins than the passive lag 2 network. The sum of the squares of the differences between the model's results and the human results are presented as a basis for comparing the model's fit to the human data.

The best ACT-R model was taken from Rutledge-Taylor and West (2005). It was exemplar based, and used the following parameters: ANS=0.28, OL=NIL. The best DSHM model was the lag 3, vector length 1024, model discussed above. The best network model was the aggressive lag 2 network.

New network versus network simulations were run for this comparison: 10000 games were run against each of the two benchmark opponents. The mean difference in wins between the aggressive lag 2 and aggressive lag 1 networks was somewhat lower than was found in West and Lebiere (2001). However, given the high standard deviation on the win differences, both the results found here and those found in West and Lebiere (2001) may be valid.

Table 2: Models of Human Rock=1 PRS

Opponent	Human	ACT-R	DSHM	Network
Agg. lag 1	9.99	12.30	10.89	5.76
Pas. lag 2	11.14	8.15	13.24	10.44
Rating		14.27	5.22	18.37

Table 2 summarizes the best models of human Rock=1 PRS play. The DSHM produces the closest fit to the human data, i.e., it scored the lowest mean squared error. Additionally, the DSHM model lost to the aggressive lag 2 model by a mean win difference of 6.22 (not shown in this table), which helps support this model as a good account of human PRS play.

Rock=2 Simulations

The mechanism for building a sensitivity to the unequal payoffs of the Rock=2 game in the DSHM models was essentially the same as for the ACT-R Rock=2 models. Three versions of the Rock=2 DSHM models were created: one that attended to opponents' rock plays more; another that attended to scissors more; and, one that gave preference to both rock and scissors. This extra attention was achieved by training the models twice on sequences of opponents' moves ending in these plays.

For each of the three variations on the Rock=2 DSHM player, the combinations of the three different lags and seven vector lengths were tested.

Evaluation Each of the DSHM players results were compared to the human data from Rutledge-Taylor and West (2004). A mean squared error approach was used. Six data point were compared: the mean points differences versus the three network models, and the strategy indices versus these opponents. Because the strategy index values were, on average smaller than the point differences, and because they are crucially important to establishing strategic play, these three data points were given twice the weight of the points difference comparisons. Thus, each model's rating was the sum of the squares of the mean point differences and twice the sum of the squares of the strategy indices.

Results The results were somewhat predictable based on the DSHM rock=2 and ACT-R Rock=2 results: The lag 3 DSHM model with vector lengths of 1024, and that paid extra attention to both rock and scissors produced the best fit to the human data. It matched five of the six data points very well. However, this model failed to defeat the aggressive lag 2 network model. There were other DSHM models that beat the aggressive lag 2 network, but failed to match the other five data points well (e.g., the margins of victory were too great).

Rock=2 Model Comparison

Three different kinds of models of human Rock=2 PRS play are discussed in this paper: ACT-R, DSHM and perceptron-like networks. The ACT-R results are taken from Rutledge-Taylor and West (2005), the DSHM model is the one discussed above. As with the Rock=1 comparison, new network versus network data was collected. The network model of human Rock=2 played each of the benchmark opponents 10000 times. Each model played 300 trial games against the three network opponents discussed in Rutledge-Taylor and West (2004). The evaluation method for all types of models was the same as for the DSHM models discussed above.

Figures 3 and 4 summarize the evaluations of the three model types. The confidence values for the human data in Figure 3 are estimated based on the 95% confidence intervals of a regression analysis of the rate of point difference achieved by the human players against each of the three opponents.

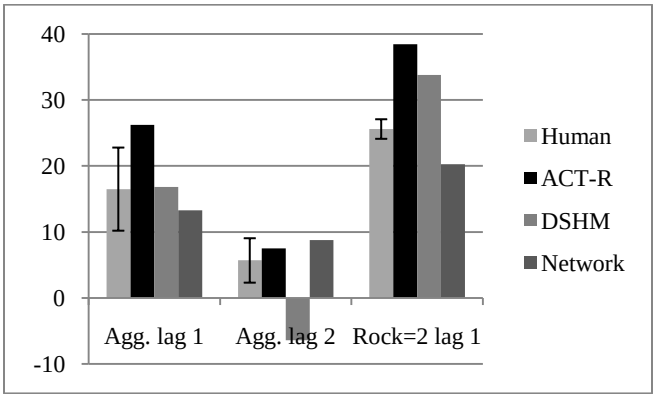

Figure 3: Points difference comparison for Rock=2 PRS. Human values include estimated 95% confidence values.

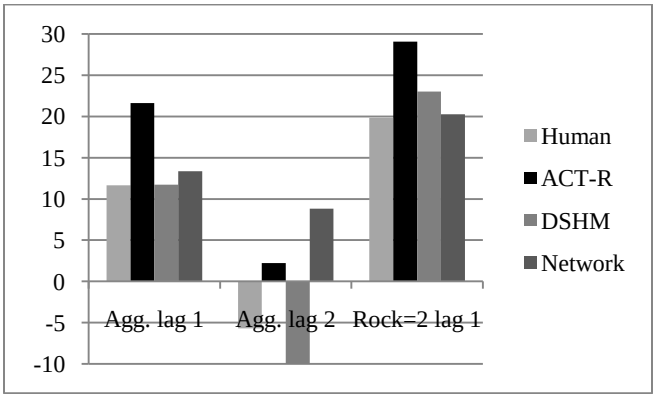

Figure 4: Strategy indices comparison for Rock=2 PRS

All three models produced good fits to the human data. However, the DSHM model is unique in that it failed to beat the aggressive lag 2 network. However, it correctly scored a negative strategy index versus this opponent. In contrast, the ACT-R and network models beat the aggressive lag 2, but did not score negative strategy indices, as did the human players. Thus, there is an obvious objective in building a superior model of human play. That is, to build a model that beats the aggressive lag 2, but does so despite a negative strategy index.

Conclusions

The simulations run and analyzed here demonstrate that DSHM can, in fact, be used successfully to model at least one memory task that relies on reconciling inconsistent information and rapidly changing predictions based on past events. Despite the fact that DSHM was designed to model

only long-term memory, it may also useful as a model of short-term memory. This also suggests that long-term and short term memory in humans may rely on the same basic mechanisms.

References

Lebiere, C. & West, R. L. (1999) A dynamic ACT-R model of simple games. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*. 296-301. Simon Fraser University: Vancouver, Canada.

Rutledge-Taylor, M. F., Pyke, A. A., West, R. L. & Lang, H. (2010) Modeling a three term fan effect. In *Proceedings of the Tenth International Conference on Cognitive Modeling*. Philadelphia, PA: Drexel University.

Rutledge-Taylor, M. F., Vellino, A. & West, R. L. (2008) A holographic associative memory recommender system. In *Proceedings of the Third International Conference on Digital Information Management*. 87-92. London, UK.

Rutledge-Taylor, M. F. & West, R. L. (2004) Cognitive modeling versus game theory: Why cognition matters. In *Proceedings of the Sixth International Conference on Cognitive Modeling*, 255-260. Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh.

Rutledge-Taylor, M. F. & West, R. L. (2005). ACT-R versus neural networks in rock=2 paper, rock, scissors. *Proceedings of the Twelfth Annual ACT-R Workshop*, 19-23. Trieste, Italy: Universita degli Studi di Trieste.

Rutledge-Taylor, M. F. & West, R. L. (2008) Modeling The fan effect using dynamically structured holographic memory. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. 385-390. Washington, DC.

West, R. L. (1998) Zero sum games as distributed cognitive systems. In *Proceedings of the Complex Games Workshop*. Tsukuba, Japan: Electrotechnical Laboratory Machine Inference Group.

West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Cognitive Systems Research, 1(4)*, 221-239.

VonNeumann, J., & Morgenstern, O. (1944) *Theory of Games and Economic Behaviour*. Princton, NJ: Princton University Press.