DEDICATION

TABLE OF CONTENTS

DEDICATION.....iii

ACKNOWLEDGMENTS .....

3.6 Performance Metrics.....108

3.6.1 Payoff.....111

3.6.2 *n*

5.7.5 Formal Conclusion from Experiment 6 215
5.8 Summary of Formal Conclusions 216
5.8 Informal Observations 220
5.9 Future Directions 221
REFERENCES 221

LIST OF FIGURES

Figure 1: A complete Collective Learning System (CLS) …….. 32

Figure 2: State Transition Matrix (STM) ……...…. 38

Figure 3: The Life Cycle of a Stimulant ……. 47

Figure 4: Sample Game States …… 92

Figure 5: Non-equivalent Game States ……... 93

Figure 6: TBL_{α} for sample states ……..... 95

Figure 7: Sample Initial and Secondary Phases …….. 97

Figure 8: All Possible Initial States ……..… 101

Figure 9: Initial States Used as Factors …… 102

Figure 10:

Figure 22:

Figure 44: Experiment 4 Results …….……..……...… 188

Figure 45: Experiment 6 Results …….……..……...… 202

Figure 46:

Table 22: Experiment 3 Results ……..……..…… 177

Table 23: Experiment 3 Results ……..……..…… 179

Table 24: Experiment 5 Results ……..……...…. 192

Table 25: Experiment 5 Results ……..……...…. 193

Table 26: Experiment 5 Results ……..……...…. 194

Table 27: Experiment 5 Results ……..……...…. 196

Table 28: Experiment 5 Results ……..……...…. 197

Table 29: Experiment 5 Results ……..……...…. 198

Defintion 27: The **secondary respondent** for a stimulant is a respondent with the second largest weight. pp 48

Defintion 28: A **supporter stim.63i4ul(n)-4(t)333]TJ /R8 12 Tf 04.8839 0 Td [( )-069.910(i)-1.9989**

Postulate 03: The

$$r_{correct}$$

Postulate 08: The compensation policy for the unordered game generates the

Postulate 09: The

Postulate 12:

Postulate 15:

LIST OF ASSUMPTIONS AND RESTRICTIONS

The fundamental ideas of Jean Piaget provide a valid psychological basis for this research. pp 12

EXECUTIVE SUMMARY

Tactic-Based Learning is an algorithm that overrides the Standard selection policy used by a Standard-CLA. A TBL-CLA follows the Standard selection policy until one stimulus is sufficiently well trained to elect its primary response as a tactic. A stimulus supports a tactic when its selection confidence is very high. Stimuli that are using a tactic (follower stimuli) simply use this response, assuming it is better than a random response. However, each follower stimulus tracks the effectiveness of the tactic and uses it only as long as it remains effective (an average compensation ≥ 1). When a new tactic becomes available, all stimuli that do not yet have an effective tactic will try it.

The lifecycle of a hypothetical stimulus in a Tactic-Based CLA is described in Figure 2. When there are no tactics in a CLA, all stimuli follow the Standard selection policy and are called seekers. As soon as the first tactic appears, all seekers will inves04308(a)4(C)-13(L)2.168

In the event that a supporter loses confidence in its response, the supporter withdraws its support from the tactic it was supporting and reverts to being independent. If the tactic no longer has any supporters, it will no longer be available for use, and any follower stimulants of it will become seekers.

failure to perform on large-scale problems. That suggestion is taken very much to heart in this research.

addition, any solution should allow the learning agent to apply previously learned

CHAPTER 2: RELATED WORK

2.1 The Psychological Basis of Infant Learning

Understanding how newborn children learn is pertinent to this research. This research is primarily concerned with learning in its earliest stages, when little or nothing has been learned about the self or the environment. The reason for this focus lies in the basic

Barr 2000), meaning that children do not need to be in the exact same situation in order to remember how to perform a task or recognize an object. These studies argue in favor of children's ability to generalize learned responses to relevant, new situations (Berk 2003).

Once children exhibit what computer scientists would call *feature-vector analysis* and what psychologists would call *generalization*, the psychology of learning and development is no longer useful for this research. Clearly, machine learning research

learning capacities, neonatal operant conditioning learning capacities, and habituation

understanding of the world around them. In this section, the stages of development are

The last phase of development is the **formal operational** stage. Most children enter this stage around 11 years of age. Whereas a concrete operational child can only operate

child. Most things fit easily into existing schemes. This is contrasted with periods of disequilibrium where a child's schemes are largely not effective for the environment and the child experiences "cognitive discomfort" (Berk 2003). When in disequilibrium, a

by operating on them in novel ways

6. *Mental representations* (18-24 months): ability to internally depict events and ideas, appearance of sudden solutions to problems

For the purposes of this research, substages 1 through 3 are of the most interest.

One of the important aspects of Dresher’s work was the ability to create new schemes, or *classes*

the *stimulus-response rules*

2.4.1 Environment

There are few requirements made on the environment in reinforcement learning. Most

to find good cell phone reception may wander around looking at the signal bars or asking

games, where a **game**

to provide feedback for *each and every* action taken by the learning agent. Therefore,

most recently chosen responses. Once an update has been calculated, it is applied to the statistics held in the STM. With the updated statistics in the STM, the CLA is ready to apply its new knowledge to the next series of stimuli presented to it by the environment.

2.5.2 Environment

The environment represents everything outside of the CLA. The environment provides stimuli to the CLA and evaluates the responses that come back from the CLA.

Definition 05: includes the necessary space, matter,

Depending on the game, the environment may or may not change its state based on

which generates an evaluation, ξ. The evaluation is usually a numeric value, but there is

interested in the effects of the compensation policy on learning (Heckman 2004), the compensation policy is usually fixed on one that seems to maximize learning during operating point pilots.

Defintion 13: The compensation policy is fixed.

2.5.7 Update

The compensation policy interprets the evaluation from the environment for the CLA. The update function then takes the compensation,

larger than the others. This is generally done by using the standard difference of two proportions to calculate the confidence that the largest probabilty is different than the other probabilities (the **reject confidence**) and then to calculate the confidence that the

3.2 Tactic-Based Learning

If an independent stimulant’s selection confidence drops below the **dependence threshold**

potencies of a reinstated tactic and the reinstated tactic must establish its global potency again.

The preceding discussion described the roles of stimulants and respondents in the

As stimulant

Since there are currently no more tactics available in the global tactic list, ϕ_k

evaluation using its compensation policy to generate a compensation value, γ. The *CLA* then passes the compensation and the history of *Responses* to the *STM*.

The bulk of the TBL algorithm is implemented in the *STM*; however, the *STM* can handle both TBL and non-TBL learning Bhand-2d9e5C1a998c89(i)-9989(c)-6(y)20(9989(or)-7.0012. 00

TrainCLA(**SEND** *initializedCLA, userInputs, truthTable, stimuli*; **RETURN** *trainedCLA*)

RETURN *response*)

TBL(

outputs, a *CLA* is evaluated based on the number of correct responses. The evaluation of a single *Response*

collection lengths, it is possible for some incorrect follower stimulants to be positively

Stimulants that make up the actual state transition matrix.

If a *CLA* is using the TBL selection policy, then the *STM* also handles the TBL process by determining when a *Stimulant* has elected or abandoned a *Tactic* and when a *Stimulant*

new *Stimulant*

As *Stimulants* withdraw their support from a *Tactic*, it is possible that a *f4*

in their responses and those that were not. A full description of the compensation policy is given in Section 3.3.4, but briefly, confident *Stimulants* should receive smaller positive updates and larger negative updates than *Stimulants* that are still learning and not yet confident in a respondent.

Postulate 09: The **update policy** for both games is

IF (ϕ is confident) THEN

$w^1 \leftarrow w$

ELSIF (ϕ*'s* tie confidence >= tie threshold AND

ϕ*'s* reject05Tf 0ct05 89(D)2.00144( )-9.92332( )]TJ ET Q 0.851563 0.851562 Tf 1.91

ID

A *Tactic* has a unique and comparable identification that is usually the same as the number of the respondent that the *Tactic* represents.

same whether the *Tactic*

ID

Boolean usedTactic

usedTactic takes that value TRUE if the *Stimulant*

Boolean TBL

TBL takes that value TRUE if the *CLA* is using the TBL selection policy and FALSE if the *CLA* is using the Standard selection policy.

3.3.11 *TruthTable*

The *TruthTable* module holds information about the domain and range sizes of the different *TruthTables*, their names, and the correct responses for each state of a given *TruthTable*

Postulate 11: The average time complexity increase for using TBL is

(a)

(b)

(a) TBL

3.4.2 Rules of the game

Initial Phase
Secondary Phase

1 target output initial TBL_{α} = 30 secondary TBL_{α} = 89,700		
2 target outputs initial TBL_{α} = 25 secondary TBL_{α} = 64,700	initial TBL_{α} =14 secondary TBL_{α} = 49,700	initial TBL_{α} =12 secondary TBL_{α} = 44,700
3 target outputs initial TBL_{α} = 12 secondary TBL_{α} = 44,700	initial TBL_{α} = 8 secondary TBL_{α} = 34,700	initial TBL_{α} = 6 secondary TBL_{α} = 29,700

1 target output

initial $\mathrm{TBL}_{\alpha} = 30$

G2.3: To measure TBL behavior when the

G3.1: To explore specific cases of the results from G2.1.

3.6.1 Payoff

At each test point, it is possible to measure the difference between the scores of the two CLAs. Because these differences are Monte Carlo averages of 30 instances of the same CLA started with different seeds for the random number generators, it is also possible to compute the confidence that the H_0

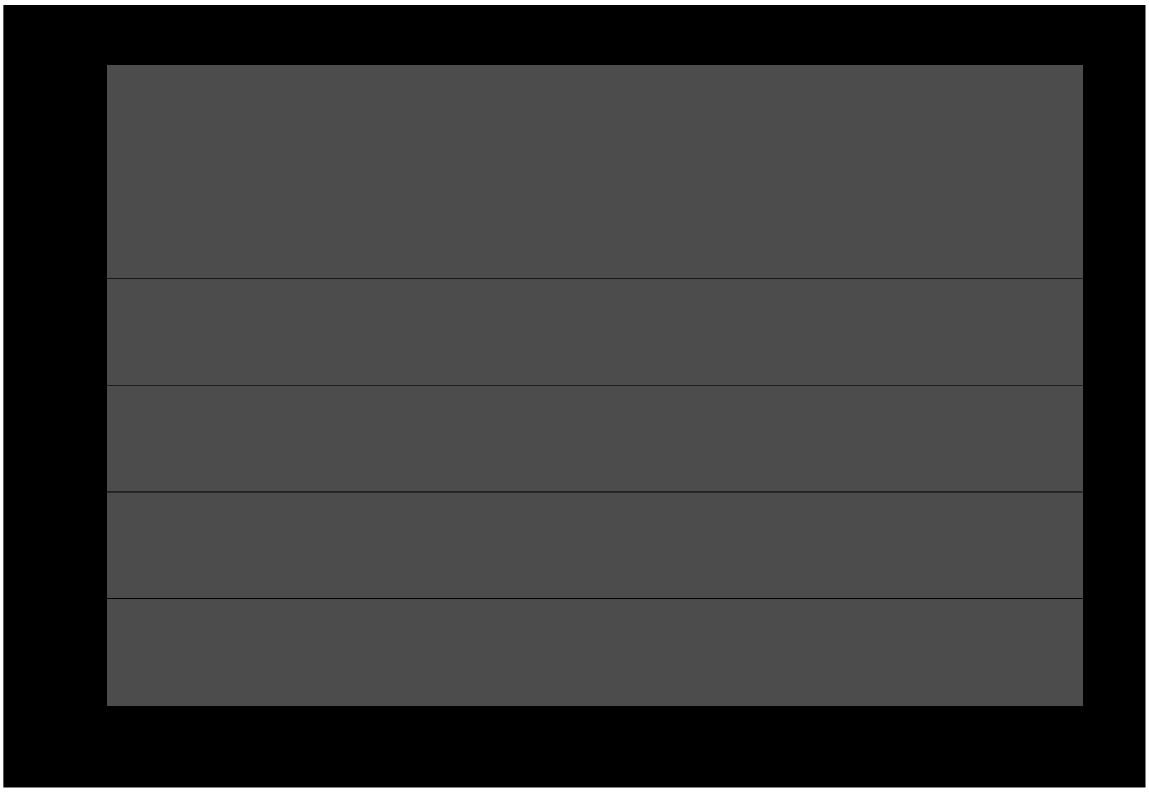
3.6.2 *n*-tile advantage

Payoff is a summary metric that assigns a single value to the entire learning process.

the distribution of the hues is not always proportional. It is important to note that the color scales are adjusted for each experiment, so care must be taken when rea

3.6.5 Learning curves

A learning curve is a graph which describes the scores of the CLA at each test point. The learning curve graph is useful for examining the behavior of the CLAs during a



test points. A contest consists of a single stimulus. If a CLA reaches 500 contests in the middle of the collection length, it completes the collection length and receives its evaluation before it begins testing. The periods between test points are known as training periods. During test points, learning is “turned off” by withholding evaluation. The CLA

4.1.3 Stationary Game, One Target Cell per Input

This section describes the informal OP Pilots necessary to determine the appropriate

4.1.3.4

□□□□□□□ below presents the block design for Experiment 4.

4.1.5 Task-switching Game, One Target Cell per Input

4.1.5.2 OP Pilot 7: TBL thresholds, Task-Switching Game, One Target Cell per Input

In order to determine the appropriate factor range, the following OP pilot was conducted in order to accomplish Goal 2.3.1 (see Section 3.5). The collection length, *c*, was fixed at 12; the compensation threshold,

- The dependence threshold must be less than or equal to the independence threshold.

$$\kappa_d \leq \kappa_i$$

4.1.5.3 OP Pilot 8: Compensation Threshold, κ_γ

The results presented in Table 19 are first sorted by minimum Payoff. There is a clear break in the distribution of minimum Payoff measures at 10.0, so the data is spilt into two groups. Each of the groups is internally sorted by the maximum Expense. The sorted data

follower to receive a very positive update one time and a very negative one the next,

5.2.2 Formal Conclusions from Experiment 1

Conclusion 2:

=70,
=70,

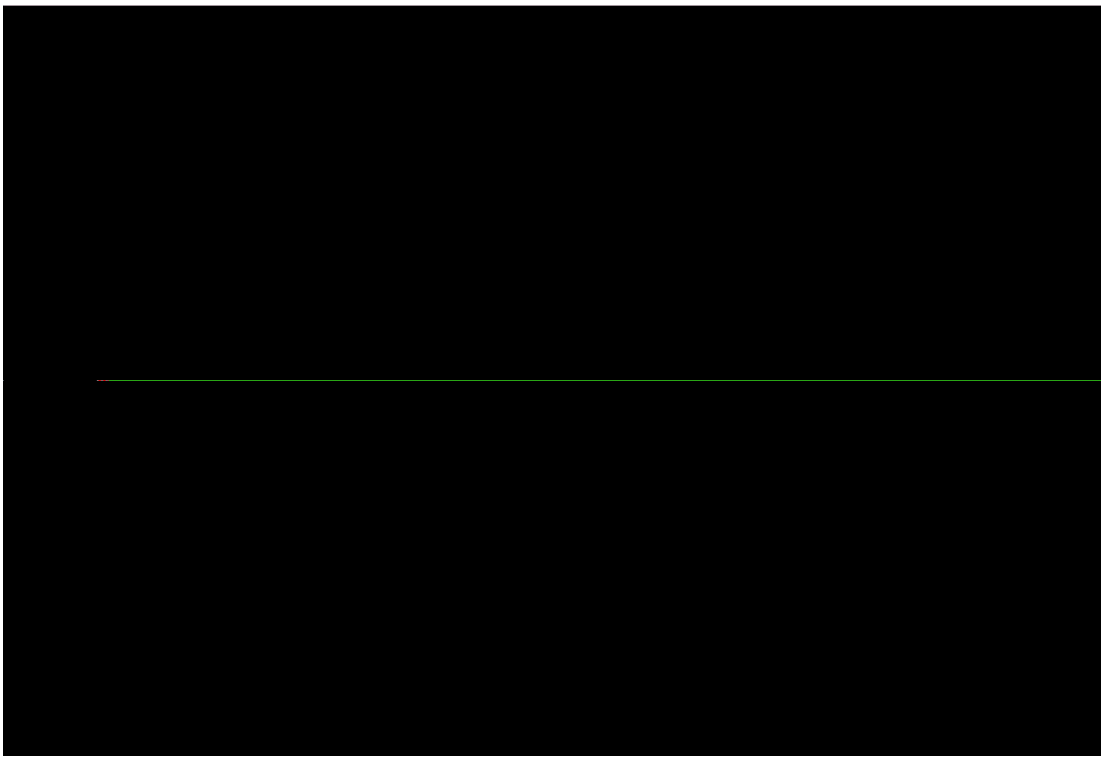
5.3.2 Best Performance: 4-tuple (κ_s=70, κ_w

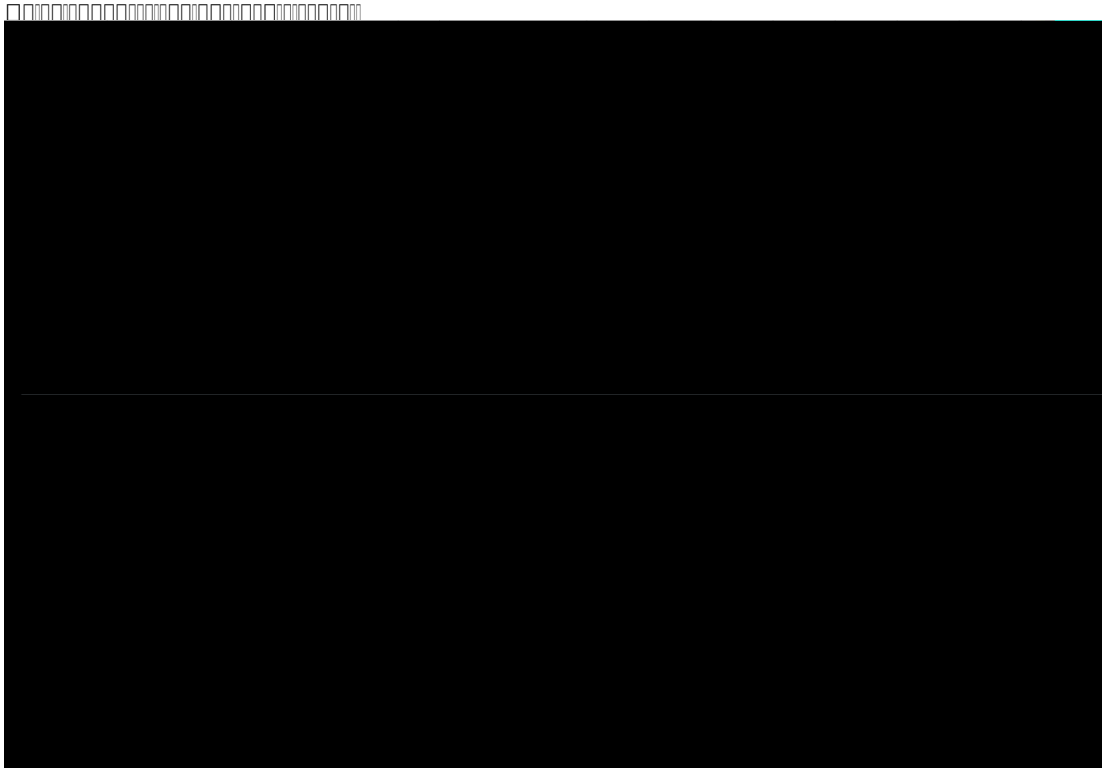
The TBL-CLA dramatically outperforms the Standard-CLA on the TruthTable game

5.3.2.2 TruthTable state 2, Collection Length 6

When the number of target responses in a TruthTable game state increases, the Tactic-Based Learning advantage, TBL_{α}, decreases (see Section 3.4.5). Figure 19 shows that as the TBL_{α} decreases, so does the performance of the TBL-CLA. When the TruthTable game state includes two target responses, the TBL-CLA still outperforms the Standard-CLA for most of the learning process, but the TBL-CLA gains, while still

5.3.2.3 TruthTable state 3, Collection Length 12



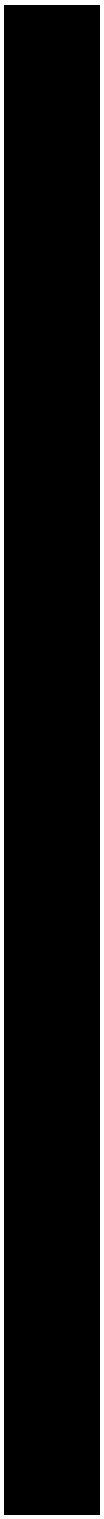


in the secondary phase of the game must be followers for at least one stage (

5.3.4.2 TruthTable state 3, Collection Length 6

While the 4-tuple shows promise at a collection length of 2 contests, it causes a TBL-

Conclusion 1d: In an environment with only one target response per input, there is no need for independent stimulants because there is no alternate response which could provide positive evaluations; therefore, a TBL-CLA performance improves when the



In order to draw more specific observations about the performance of a TBL-CLA, it is necessary to reorder the footprint and examine smaller subsections of it. In order to

APPENDIX H: EXPERIMENT 5, EXPENSE RESULTS; and APPENDIX I: EXPERIMENT 5 *N*-TILE RESULTS.

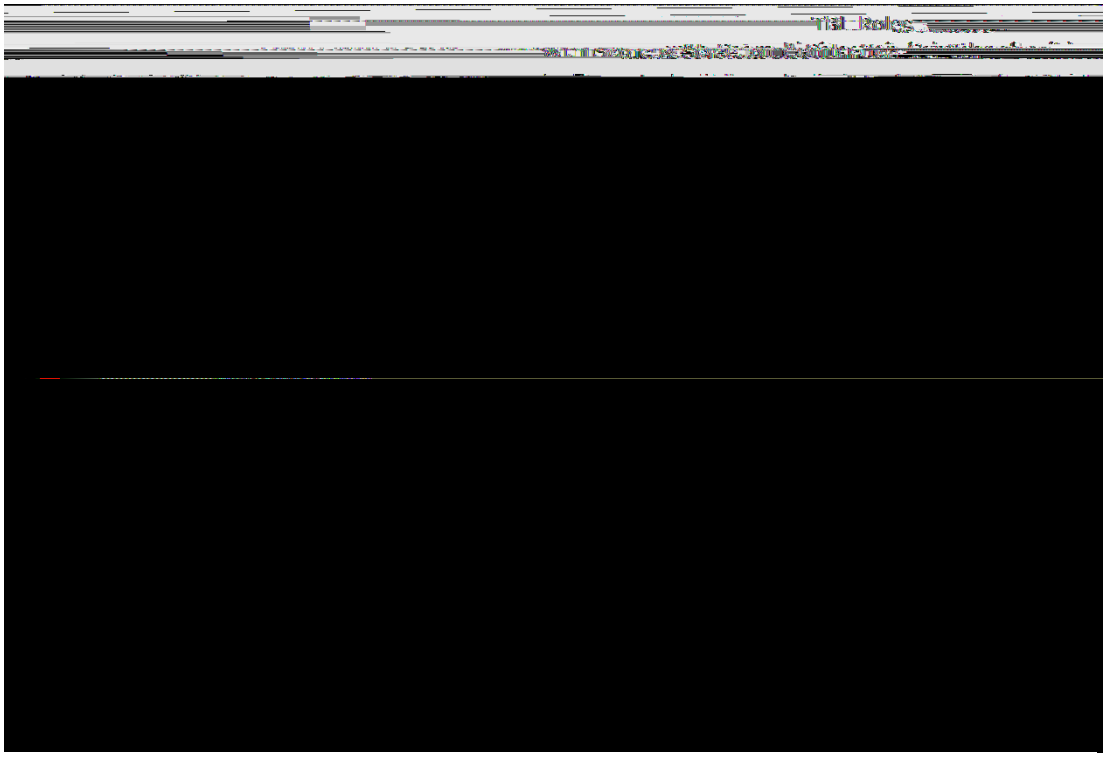
Table 24 shows the results of Experiment 5 sorted by the TBL thresholds. Section 3.6.4 describes the layout of the footprint in greater detail, but briefly, the columns and rows are organized in a hierarchical fashion.

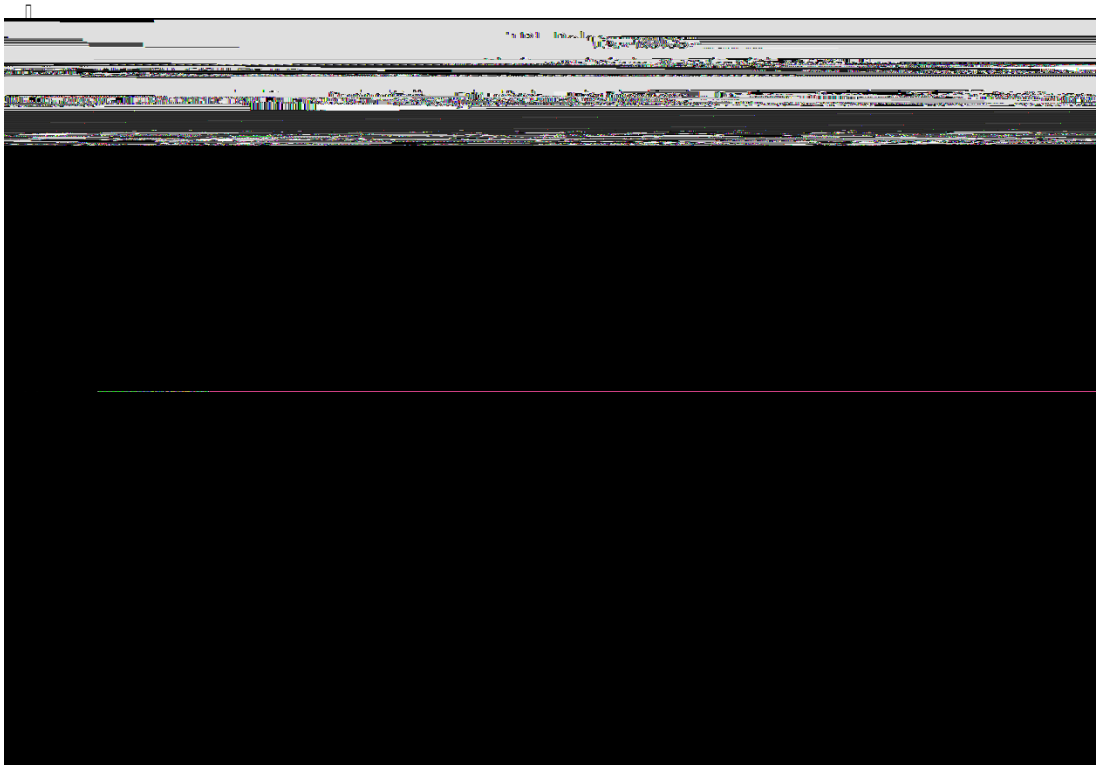
The columns are divided first by the performance metric: Payoff, then Expense, then *n*-tile advantage. Within each performance metric, the columns are subdivided by collection length. Finally, within each collection length, the columns are again subdivided

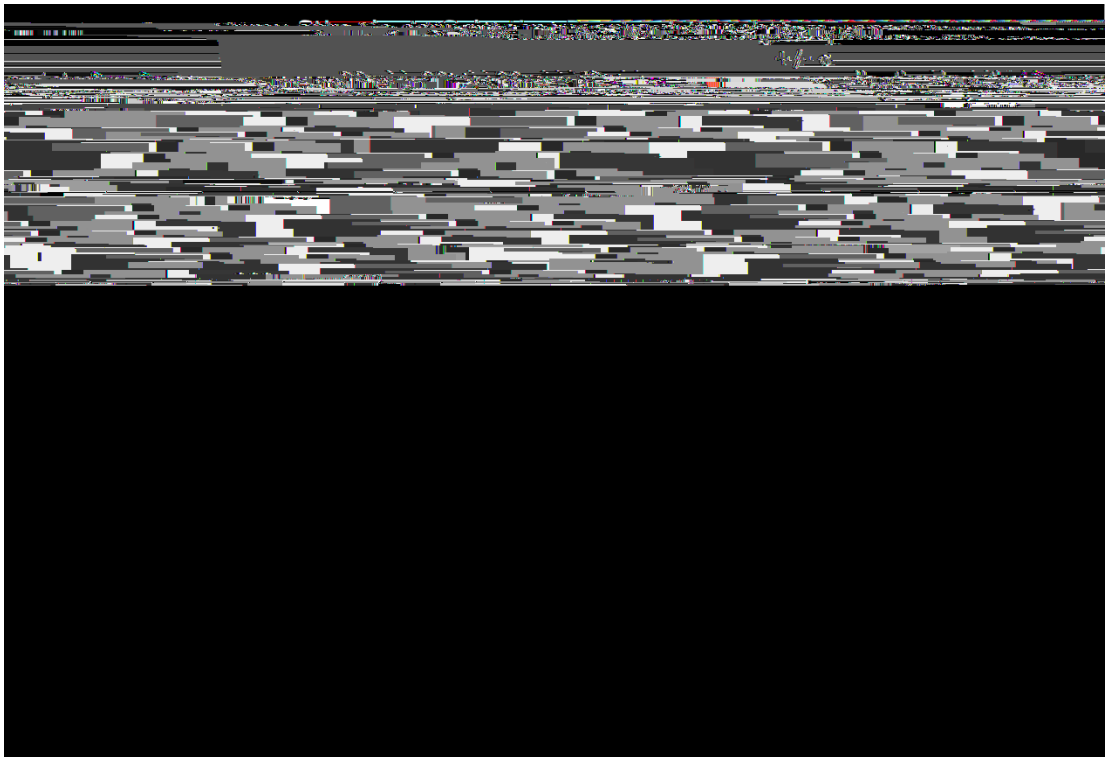
- The dependence threshold must be less than or equal to the independence

Turning to the results in Table 24, it can be seen that the TBL-CLA does not have as strong as an advantage in a task-switching game as it is in the other tw

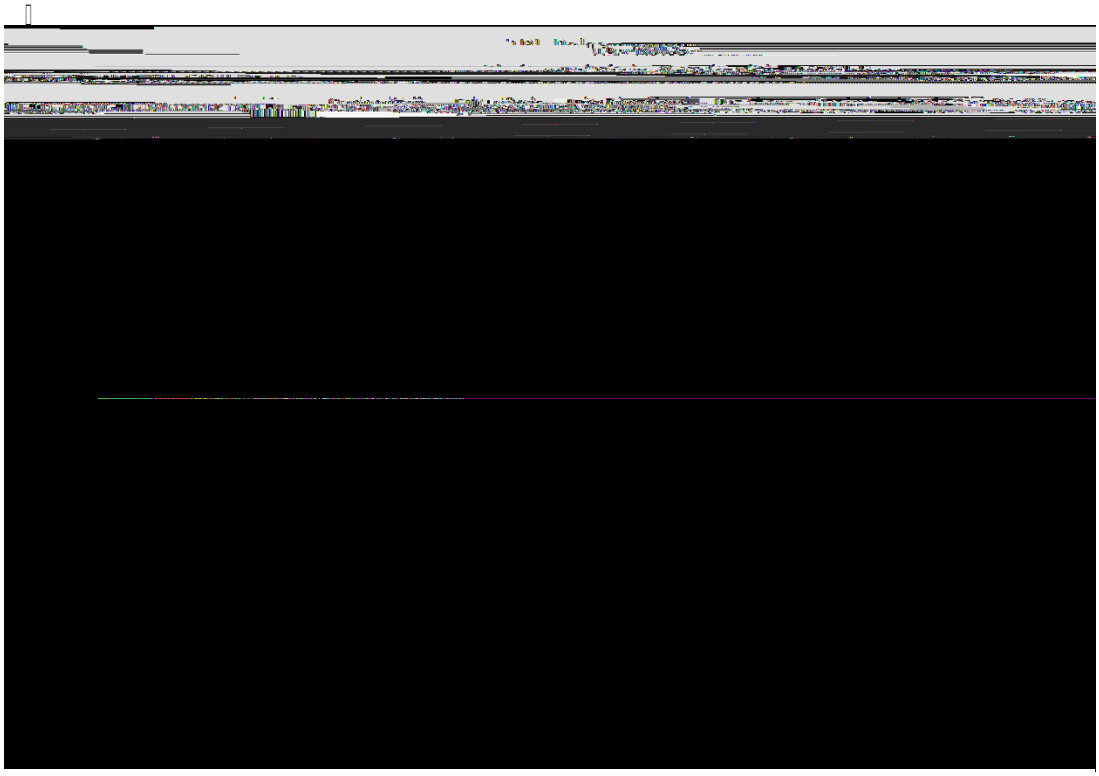
5.7 Experiment 6





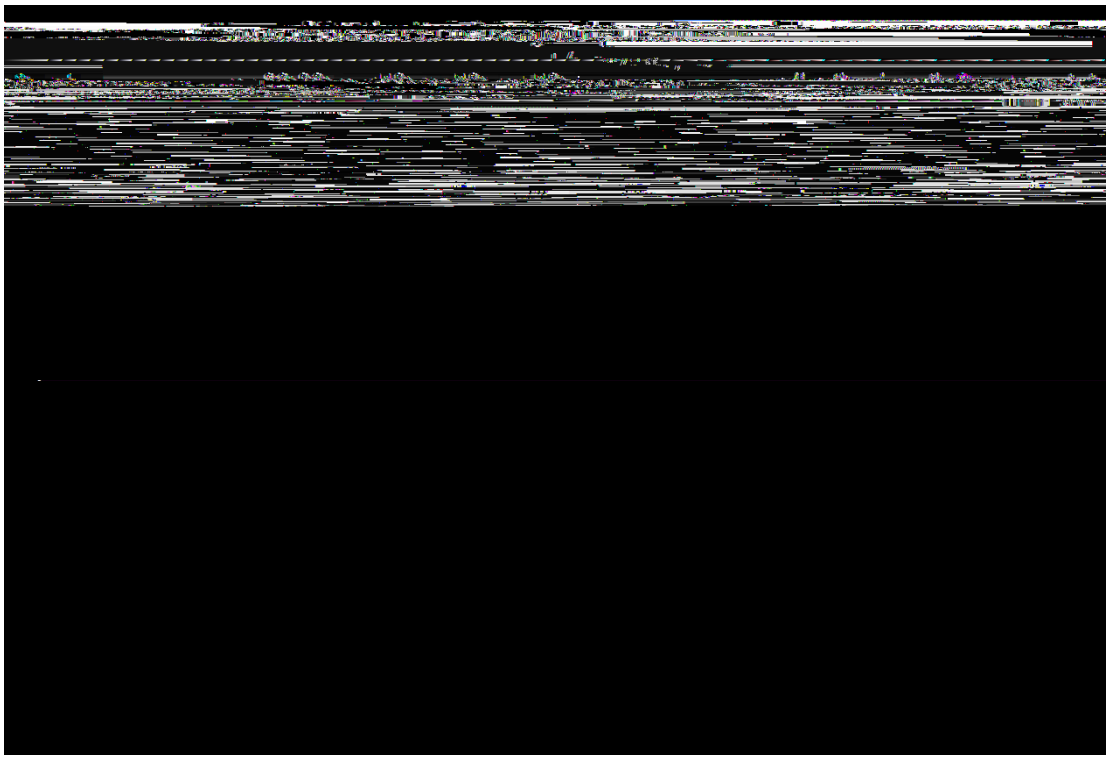


5.7.2.3



5.7.3

5.7.4 Poor Performance: 4-tuple (



5.8 Summary of Formal Conclusions

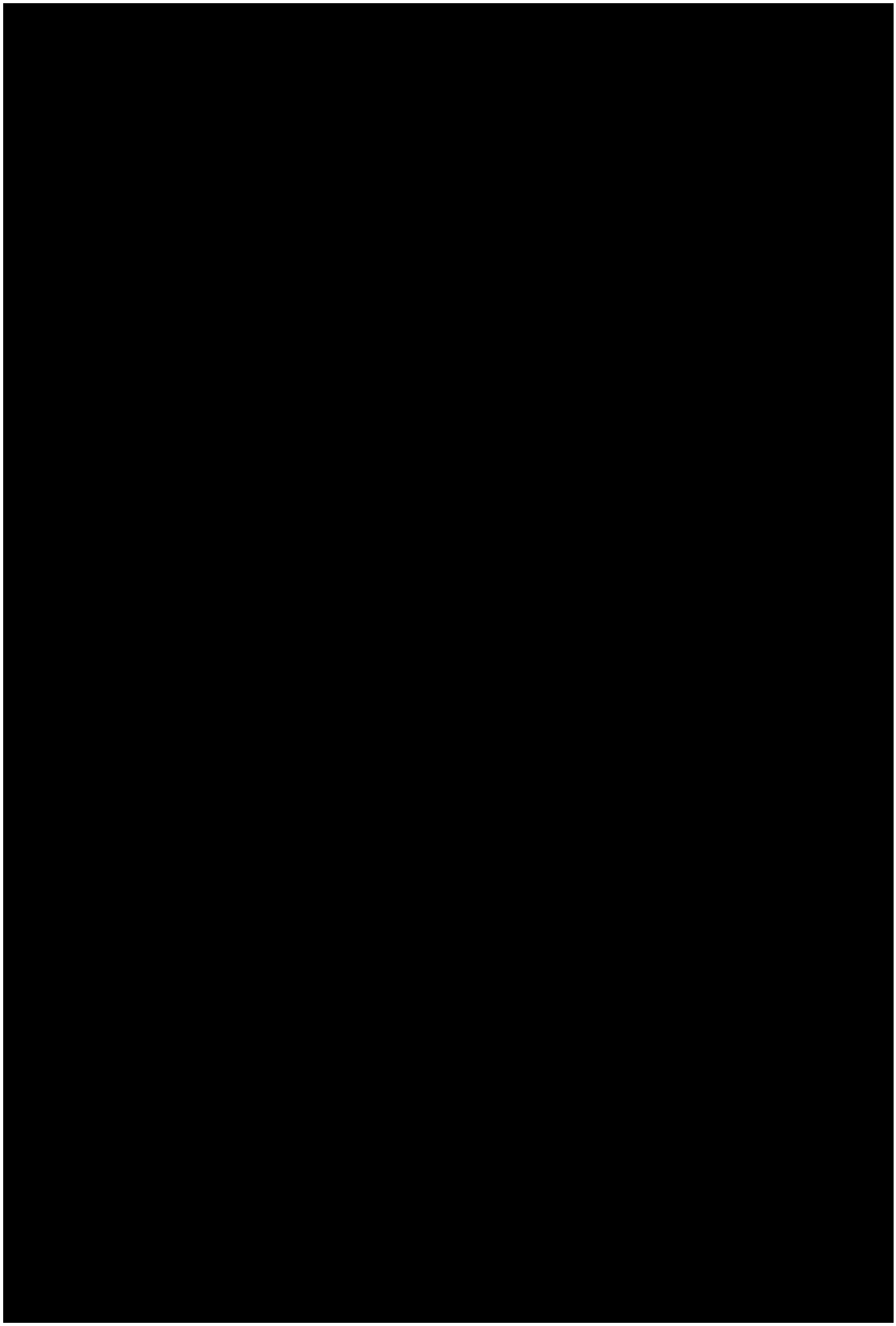
Conclusion 5b:

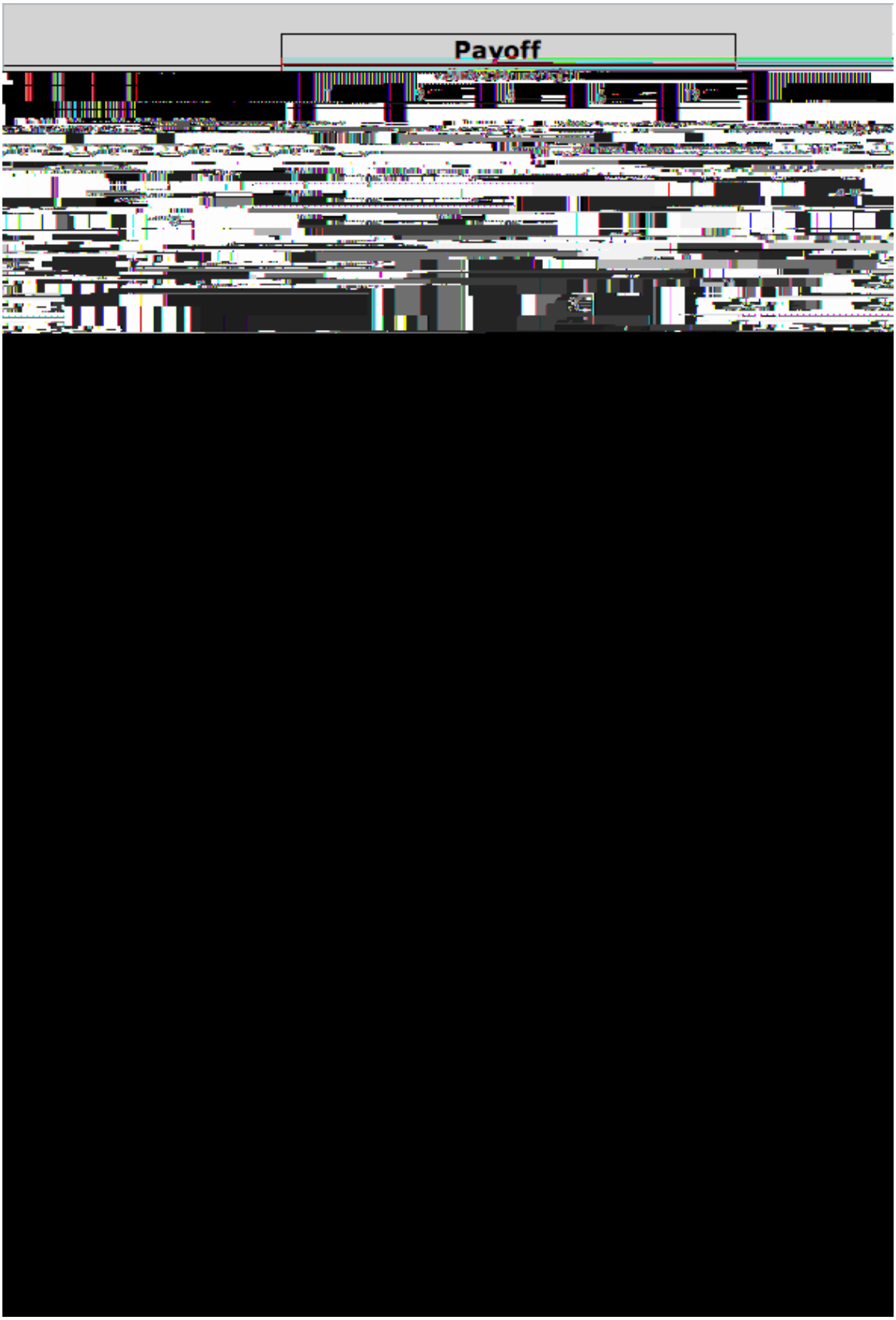
responses are equally correct, so once a stimulant has found one target cell, there

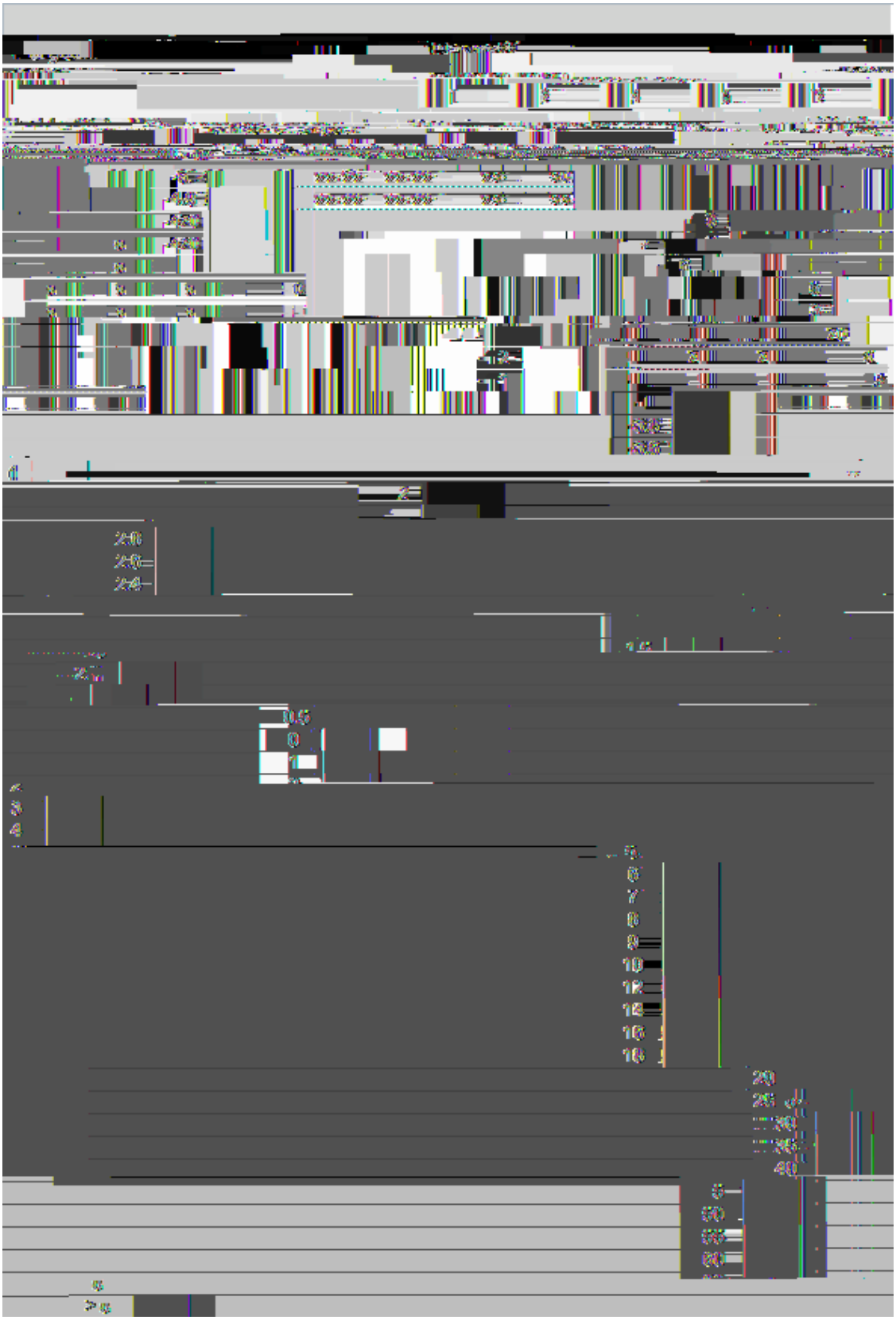
REFERENCES

Bock, P., (1993) *The Emergence of Artificial Cognition: An Introduction to Collective Learning*, World Scientific, New Jersey.

Bock, P., Rovener, R., Kocinski, C. J. (1990) “A Performance Evaluation of ALIAS for the




Payoff




Expense