# NEURAL NETWORKS AS A LEARNING PARADIGM FOR GENERAL NORMAL FORM GAMES[1]

LEONIDAS SPILIOPOULOS[2,3]

Abstract

This paper addresses how neural networks learn to play one-shot normal form games through experience in an environment of randomly generated game payoffs and randomly selected opponents. This agent based computational approach allows the modeling of learning all strategic types of normal form games, irregardless of the number of pure and mixed strategy Nash equilibria that they exhibit. This is a more realistic model of learning than the oft used models in the game theory learning literature which are usually restricted either to repeated games against the same opponent (or games with different payoffs but belonging to the same strategic class). The neural network agents were found to approximate human behavior in experimental one-shot games very well as the Spearman correlation coefficients between their behavior and that of human subjects ranged from 0.49 to 0.8857 across numerous experimental studies. Also, they exhibited the endogenous emergence of heuristics that have been found effective in describing human behavior in one-shot games. The notion of bounded rationality is explored by varying the topologies of the neural networks, which indirectly affects their ability to act as universal approximators of any function. The neural networks' behavior was assessed across various dimensions such as convergence to Nash equilibria, equilibrium selection and adherence to principles of iterated dominance.

KEYWORDS: Behavioral game theory; Learning; Global games; Neural networks; Agent-based computational economics; Simulations; Complex adaptive systems; Artificial intelligence

.

## 1. INTRODUCTION

There are two main criticisms that can be levied against the main research directions of the game theory learning literature. Firstly, that most research involves repeated games against the same opponent and secondly that fixed learning rules are often employed which are ad hoc specifications of the researcher. These tactics are primarily ones of convenience, as studying learning in the same game ignores the complications of transfer of learning across different yet similar games, and assuming that a player is facing the same opponent greatly simplifies learning dynamics. Many learning rules cannot be applied to a series of one-shot games because they are specifically formulated to explain behavior only when repeated instances of the same game are given to subjects. For example, reinforcement learning and fictitious play assign values over time to players' actions, however this approach only makes sense if the payoffs of the game do not change over

[1]This is the one of three papers towards a PhD degree, conferred by the University of Sydney in May 2008.

[2]University of Sydney

[3]The author would like to thank Kunal Sengupta for constructive input and thesis supervision. This research was made possible by funding provided by the Government Department of Education, Science and Training (Australian Government), the College of Humanities and Social Sciences and the Faculty of Economics and Business at the University of Sydney.

time. Otherwise, the actions will be qualitatively different and will not be able to be updated according to a unique, time-stationary equation.

The previous paper in this thesis loosened to a certain degree the assumption of repeatedly playing a single game against the same player by examining learning of repeated games across three different computer algorithms, leading to insights regarding the separation of game-specific and opponent-specific learning which is otherwise obfuscated. This paper now proceeds to fully abandon the assumptions of repeated game play with the same opponent in an effort to examine learning over a wide variety of games, with very different strategic structures and randomly rematched antagonists.

Learning in an environment with random rematching of players and random draws of games from a strategically diverse set of games is closely linked to the experimental literature on one-shot games. This is exactly the kind of learning experience that a subject would have faced prior to participating in an experiment and will likely draw upon when deciding how to play a game never seen before. The experimental literature on one-shot games is primarily concerned with uncovering the decision rules used by subjects when faced with a game for the first time. The standard approach in this field is for researchers to postulate what they regard to be reasonable rules and check for evidence of their use by human subjects. This approach does not really study any type of learning mechanism as it simply assumes people come into the experiment with some experience and then play each game once. It must be conceded that this is an inherent limitation of the experimental process for it is not possible, at least within ethical bounds, to access the knowledge and experience human subjects bring with them to an experiment. The rules that researchers detect in one-shot game experiments have been learned by subjects prior to entering the laboratory over a long period of time from everyday encounters in their lives, where they have faced very different games and opponents. This limitation of the experimental method is what has essentially restricted progress in explaining the formation of observed rules and therefore what little research has been done in this field has been theoretical.

Several standard learning theories have been modified to be amenable to stochastic analysis allowing for some results to be deduced regarding convergence to Nash equilibria and other behavior. As will be discussed in Section 2, these theoretical approaches require quite restrictive simplifying assumptions in order to make the analysis tractable.

This paper contends that simulations of neural network agents[4] in an agent-based computational economics

[4]Despite their widespread use in the field of psychology and cognitive neuroscience as models of the human mind, their use as models of economic agents has been fairly limited. Notable exceptions to this include Cho and Sargent (1996) who strongly support the use of NNs as models of behavior. They use perceptrons, or simple NNs, as a way of modeling bounded rational behavior in repeated situations, such as a repeated prisoner's dilemma. They show that even the simplest NN, a single perceptron, is capable of implementing trigger strategies in a repeated prisoner's dilemma game and of supporting all subgame perfect payoffs. They also prove that only a slightly more complicated network is capable of supporting all equilibrium payoffs in a general $2 \times 2$ game. Sargent (1993) models economic agents as simple NNs and looks at the resulting macroeconomic dynamics of an economy populated by such agents. Most applications of NNs in economics, business and finance have been in the field of forecasting and prediction, such as exchange rates (Kuan and Liu, 1995), stock price movements (Leung et al., 2000), inflation forecasting (Nakamura, 2005) and bankruptcy (Yang et al., 1999). For a

model are the ideal way to circumvent many of the problems related to the non-tractability of theoretical models, thereby effectively complementing such research. The universal approximation theorem (Hornik, 1991; Cybenko, 1989; Funahashi, 1989) states that given enough neurons a single layer neural network (NN) is a universal function approximator i.e. it will be able to approximate any function, with any desired level of accuracy[5]. Hence, using a NN model evades the problem of knowing and specifying the exact functional form of the learning rules employed by humans. On the contrary, the final functional form arises through the learning of the NN and the adjusting of its parameters - this can be thought of as a type of "learning to learn". This is in stark contrast to the existing literature where specific functional forms are posited by the researcher and then tested e.g. fictitious play, reinforcement learning, EWA. Hence, using a NN to model learning can be thought of as a kind of endogenous general to specific approach, since specific rules are not burned into the model but rather evolve endogenously starting from a very general model and guided by a very basic principle of adjusting its parameters according to myopic best response to the previous outcome.

There are many advantages to such an approach. Firstly, such a setup will provide a learning model of one-shot play rather than simply positing various heuristics[6] to explain ex post the observed one-shot play of humans. Secondly, the learning rules that will arise in the NNs can be used to deduce the plausibility and appropriateness of the rules often postulated by researchers in the one-shot game literature and perhaps propose new rules.

The literature has employed very different specific learning rules according to the type of game being played. The main reason for this is that either the models used can by definition only be applied to specific games, or perhaps the same rules do not perform well on different games. That learning models perform well in some cases and not others simply implies that the true underlying laws governing behavior have not been discovered, only rough approximations to behavior which is why different models need to be postulated.

A solution to this problem is to use a learning model that maps the payoffs of a game directly to probabilities of playing actions. Standard statistical models with prespecified functional forms can map payoffs to actions however there are important reasons to prefer NNs instead of standard models, as will be defended in Section 3.1.

The stratagem of this paper is to present in Section 2 the relevant theoretical literature in regards to the reasoning behind the choice of neural networks as representatives of learning agents in simulations. Also, the theoretical literature on learning in global games will be reviewed with emphasis on how simulations can complement the theoretical models by simulating more realistic models that are too complex to be analyzed

collection of papers on neural networks applications the reader is referred to Smith and Gupta (2002).

[5]This result however is conditional on the ability to reach the global minimum, which is dependent on the complexity of the error surface and the training rule used to obtain the weights of the NN.

[6]A detailed description of the most commonly used statistics is given in Section 5.7.

theoretically. An in-depth discussion of the paper by Sgroi and Zizzo (2002) is carried out, followed by specific proposals to expanding this model. An informal introduction to neural networks and the methodology and specification of the simulations follows in Section 3.3. A brief discussion of the results obtained for $2 \times 2$ games is followed by an in depth analysis for $3 \times 3$ games in Sections 4 and 5. Specific issues that will be broached include the convergence (or lack thereof) to Nash equilibria, play of dominated and iterated dominated strategies, payoff comparisons of different NN topologies, equilibrium selection in relation to risk- and payoff-dominant equilibria, inference of simple heuristics that have endogenously emerged as a result of learning and a discussion of the dynamic behavior of the NNs during the training process. All the while, comparisons of NN behavior on the above dimensions will be made to human subject experiments of one-shot games in an effort to ascertain whether NNs are a plausible learning model for general normal form games. Finally, Section 7 discusses the various conclusions and avenues for future research. Appendices include technical discussions of neural networks and the backpropagation algorithm, and results from other simulations not discussed in the main text.

## 2. LITERATURE REVIEW

### 2.1. *Learning in global games*

The past few years have seen the emergence of some theoretical work into generalizing learning rules so that they are applicable to a wider range of games. However, this literature is limited as a consequence of the methodology of attempting to derive analytical and mathematical proofs of the properties of these learning systems. There are three problematic assumptions which are used to make the models tractable to analytics.

The first simplification is to use an exogenously postulated similarity measure with the intent of basing a decision on a newly presented game on past experience with similar games. This case-based approach to learning games is adopted in Steiner and Stewart (2006).

A second means of simplifying analysis, followed by Germano (2007) reverts to postulating rules, which by construct, can prescribe play for any arbitrary $3 \times 3$ game and subjects these rules to evolutionary selection, where the probability of each rule being used depends on it past performance. One of the main results he finds is that any rules which do not survive iterated elimination of dominated strategies in the average game will tend to disappear in the long run. However, as is acknowledged, the approach is hindered by the necessity of the researcher to postulate the set of learning rules that will be part of the initial population to be subjected to evolution. An experimental paper by Stahl and Haruvy (2004) has subjects play a sequence of 30 different symmetric $3 \times 3$ games under the assumption they are learning to use a set of exogenously postulated rules, which are essentially best response rules of varying depth. They find that over time the depth of best response increases and that most subjects are implicitly assuming that other players are also increasing in sophistication.

Finally, a third approach restricts analysis to games which may have different payoffs but common strategic forms. LiCalzi (1995) similarly discusses fictitious play learning in classes of games that are predetermined to be strategically similar and specifically states that although it would be interesting to examine how similarity and learning could be modeled in games with different strategic forms, the technical difficulties involved with proving theorems analytically is daunting.

Katz (1996) in her Ph.D. thesis discusses the importance of allowing similarity measures to be modeled as a learning process just as beliefs are. A general extension of learning in $2 \times 2$ normal form games is presented where from a theoretical aspect each game is preceded by a stage where the player chooses how to partition past games according to a similarity measure. She discusses some of the general implications in repeated games of partitioning the game set but does not provide or analyze specific models of similarity learning.

All of the above research is forced to use at least one of the three simplifying assumptions in order to attain proofs regarding the convergence properties of these learning models. Ideally, none of these three assumptions should be employed however the general consensus is that this is extremely difficult to accomplish for analytical proofs. The solution to this problem is to change methodology and turn instead to the use of simulations, which will be able to answer and provide insight into more complex and realistic models of learning a broad class of games. The paper discussed below in detail is an example of such a methodology, whose analysis however is restricted only to games with a unique PSNE. However, proposed modifications to this line of research will allow for learning of any type of $n \times n$ game, thereby achieving the research goals without the use of these three unrealistic simplifying assumptions.

### 2.2. *Strategy learning in $3 \times 3$ games by NNs (Sgroi and Zizzo, 2002)*

The closest line of research to this paper is Sgroi and Zizzo (2002), and an abridged published version of the paper (Sgroi and Zizzo, 2007), henceforth collectively referred to as S&Z. In their research, they used a set of $3 \times 3$ games whose payoffs were randomly sampled from a uniform distribution with the restriction that the games had only one pure strategy NE. A single NN was released into a general population of Nash equilibrium players and was randomly rematched to play against them. Hence, after each round of play the NN would attempt to learn the Nash equilibrium of each game by altering its weights according to a backpropagation rule, which is a standard NN learning rule. The goal was to determine how successful the NN would be in identifying the NE strategy in games it had never seen before. Ideally, after presentation of many different games the NN will have learned how to generalize from these games and therefore will be able to pick out the NE strategy with better than random performance.

The main findings of this paper are the following.

1. The trained NN was able to select the Nash equilibrium action in games it had never encountered before at a rate of 60%-80% of the time. The difference in these rates is attributed to the definition of a correct response by the neural network. If the neural network's response has to be within 0.05 of the ideal response, then the PSNE is chosen roughly 60% of the time, whereas if the loosest criterion is employed so that the NNs' response has to be within 0.5 of the ideal response then a robust upper limit of 80% is achieved.
2. The responses of the trained NN could be approximated quite well by simple heuristics or bounded rational rules that emerged endogenously due to the learning process of the NN. The heuristic which best described the behavior of the trained NN was the $L2$ heuristic which simply best responds to the assumption that one's opponent uses the $L1$ heuristic. This heuristic assumes that one's opponent will play each action with equal probability and then best responds to this. The $L2$ heuristic can predict the chosen action of the NN 85% of the time.
3. If the trained NNs are presented with games with multiple NE they tend to exhibit focal points in the sense that different networks tend to converge to the same equilibria. This occurs even if they are trained on a different set of games, albeit with the same game sampling scheme. From these results they infer that the NNs must have learned to detect strategic features of the games they faced which led them to the same solution in games with multiple PSNE.

## 3. METHODOLOGY

### 3.1. *In defense of the use of NNs as computational learning agents*

This section presents the rationale behind the choice of NNs, which are quite complex models of cognition and intelligence, over other candidates to model human learning and behavior in strategic games.

*Imposition of minimal prior assumptions including no exogenously imposed concept of similarity of games*

This argument refers to the fact that the researcher need specify a minimal amount of functional form or structure for NNs compared to the usual learning rules posited in the literature. An obvious approach for a set of very different games is to postulate a learning process for each possible game in this set. This is only feasible however if there is a small number of games in this set, a significant restriction especially given that there could exist a very large, or even infinite, number of games. A better approach is to break games down into a much smaller number of classes by using a similarity measure to group games together - this is essentially a case-based reasoning model (CBR), see Gilboa and Schmeidler (1995) and Leake (1996). Different learning rules could then be applied to each group or class of games. These rules could have the same functional form but their own unique set of updatable parameters or could even have completely different functional forms.

Either way, the important concept to bear in mind is that the learning model for each similarity group will be independent of all the other groups. This approach is problematic because the researcher is forced to define a similarity measure with which to group the games, rather than allow these classes or similarity measures to be learned endogenously.

*A more biologically realistic implementation of learning*

The use of neural networks is an important step towards biological plausibility of learning processes because it utilizes a distributed knowledge topology as does the human brain. Although specific information on how the connections between the neurons in the brain are adapted is only now beginning to emerge, the relative simplicity of the backpropagation algorithm compared to other algorithms used in the NN literature and its popularity make it a desirable choice. Other more powerful algorithms exist however their use of second-order information seems highly biologically implausible, whereas the simple first-order gradient descent technique used in backpropagation is highly plausible, even though the specific calculations employed in the algorithm may not be exactly the same as those in the human mind. The use of the backpropagation learning algorithm as approximating real neuronal adaptation in the human mind is defended by Robinson (2000), Zipser and Andersen (1988), Mazzoni et al. (1991), Kettner et al. (1993), Lehky and Sejnowski (1988) and Dror and Gallogly (1999).

*The nesting of many other types of models as special cases of NNs*

Special topologies of NNs exhibit simple analogues to standard statistical models whilst still allowing for a dynamic learning procedure rather than an instantaneous closed form solution. For example, by varying the topology of a NN it is possible to emulate the following standard econometric models: logistic regression (Hosmer and Lemeshow, 1989), multivariate multiple-linear regression (Myers, 1986) and polynomial regression (Pao, 1989)[7].

*Minimal memory requirements*

Assuming that the history of play will at some point become quite large the issue of memory storage, access and search becomes important. NNs are an extremely desirable solution to this predicament as they essentially automatically perform information compression. The only parameters which need updating in NNs are the weights and biases which will consist of much fewer bits of information than perfect storage of the information from all games played. Therefore, the parameters of the NNs embody the complete past history of games played albeit in an imperfect manner. A desirable property of NNs is that the amount of memory

[7]For a more general discussion of the similarities between statistical models and NNs the reader is referred to Sarle (1994).

or information storage is independent of the size of experience or past history. The problem of accessing and searching through a large database of prior memories is solved since no searching is necessary.

### 3.2. *Proposed modifications to Sgroi and Zizzo (2002)*

1. In S&Z, the NNs were pitted against other agents that played the NE every single time. This setup is a type of supervised learning where the NN is told at the end of each game what the NE strategy was and by implication what the correct answer or desired output was. This however only tests the computational ability of NNs to learn the NE when explicitly being trained for this purpose. A more realistic setup is to use an Agent-based Computational Economics[8] (ACE) approach to this problem, where a population of NN agents are learning to play a set of games concurrently with zero initial experience, based on their local interactions with each other (with no knowledge of a NE assumed or imposed on any of the agents). Even if a NN does have the capability to learn the NE when trained by presenting it with the correct result every time, this does not necessarily imply that a population of NNs learning simultaneously from each other will identify the NE without an external teacher directing their learning.
2. The desired output in S&Z for the NN was the unique PSNE action. This paper follows a different tack where each NN will observe its opponent's realized action and will assume that myopic best response conditional on its opponent's realized action is the desired output. Hence, if a NN's opponent is not playing according to the NE then it has no incentive to play the NE either, but instead must figure out a strategy that best responds to its opponent's strategy. This is a much looser and realistic assumption as it does not require the existence of an "external teacher" to tell the NNs what the correct NE action is. If Nash equilibria are learned in such a setup it will be due to procedural rationality arising from the interactions of the agents and their attempts to better respond to their environment. An important observation is that the desired output of a specific game in this case depends on other agents' actions which are changing over time due to learning, whereas in Sgroi and Zizzo (2002) the desired output will always be the NE solution. This formulation of desired output now allows the training of NNs on all possible games without restriction to games with a unique PSNE. This restriction was necessary in S&Z due to their setup where NNs where given the "correct" answer, implying that there cannot be

[8]Multi-agent modeling and its application to economics, coined as Agent-based Computational Economics (ACE), is an interdisciplinary field at the borders of evolutionary economics, cognitive science and computer science. Waldrop (1992) attributes the following definition of a CAS to John H. Holland - "A Complex Adaptive System (CAS) is a dynamic network of many agents (which may represent cells, species, individuals, firms, nations) acting in parallel, constantly acting and reacting to what the other agents are doing. The control of a CAS tends to be highly dispersed and decentralized. If there is to be any coherent behavior in the system, it has to arise from competition and cooperation among the agents themselves. The overall behavior of the system is the result of a huge number of decisions made every moment by many individual agents". For an introduction to the literature the reader is referred to (Tesfatsion, 2002) and (Tesfatsion and Judd, 2006).

more than one such answer or that this answer cannot be changing over time for a given game.

3. The output of the NNs in S&Z was not designed to be interpretable as a probability distribution over a NN's action space. This immediately precludes the use of these networks for games with a unique MSNE, and does not adhere to the experimental evidence that humans employ stochastic rather than deterministic decision rules. The topology of this paper's proposed NNs is such that they directly output a probability of playing each action by using a logistic or Boltzmann distribution in the output layer neurons, thereby providing the necessary framework for learning general normal form games. Hence, the trained NNs will be able to provide insight into equilibrium selection in games with multiple PSNE, such as coordination games, or games with a unique mixed strategy Nash equilibrium.
4. An attempt to extract rules from the NN will be made to infer what the networks have learned. NN output will be regressed on the heuristics that human subjects have been found to depend on in the experimental literature of one-shot games, as a means of determining whether they are similar. Finally, NN output will be directly compared to the behavior of human subjects in many experimental studies of $3 \times 3$ one-shot games.
5. Another variation implemented in this paper is the use of online or incremental learning instead of batch learning that was used in S&Z. With online learning, NN parameters are updated after the presentation of each observation, consisting of the input and desired output of one game, whereas with batch learning they are updated after one pass of the whole set of observations. For example, with batch training all the observations in the training set would be fed to the NNs and then the parameters would be updated based on the errors from all of these observations together. The networks will be repeatedly presented the whole set of observations with updating occurring at the end of each presentation, and the process continuing until a prespecified stopping point. With online learning, weights are adjusted after the presentation of each observation, which is clearly more ecologically valid or realistic than adjusting weights after the whole set of observations. Most of the available learning algorithms in the literature can be used for batch training with only a much smaller subset available for online learning. This occurs because many algorithms adapt their own parameters after the presentation of inputs. It is possible to do this with batch training where learning occurs after the exact same set of observations has been presented, but in online learning it would not make sense to compare results from the previous observations as they are completely different from the current observation. The three advantages of online learning (also referred to as stochastic learning or noisy learning in the NN literature) according to LeCun et al. (1998) are the following. Stochastic learning leads to faster convergence, better solutions in the sense of avoiding local minima and is better suited to tracking changes in the environment if it is non-stationary. Stochastic learning obliges gradient descent to be applied in a noisy fashion so that it is

more likely to avoid becoming stuck in bad local minima as it will tend to explore a larger area of the error surface than batch learning. Tracking changes over time is very important to this application as the correct output in each case depends on the behavior of other NNs which are learning simultaneously, implying that their behavior will be changing over time (i.e. that the correct outputs for given games are non-stationary over time).

### 3.3. *An informal introduction to neural networks*

A brief introduction to neural networks follows with Figure 1 providing a graphical display of the structure of a NN, whilst a technical treatment of NNs is reserved for Appendix A.1.

A NN consists of layers of interconnected simple processing units called neurons, whose functioning is based on that of real biological neurons. Each neuron receives a range of inputs which are summed according to weight parameters, upon which it then performs a non-linear mapping or transformation and finally outputs the resulting value.

There exist three distinct types of layers of neurons. The first, or input layer, receives input from the environment (defined as anything outside the NN structure), the hidden layer(s) internally process this information and have no contact whatsoever with the environment, and finally the output layer submits the output of the network to the environment. In this application, the input from the environment is simply the payoff matrix of the normal form game being played and the number of neurons in this layer corresponds to the sum of the sizes of the payoff spaces of the two players. Thus, for $2 \times 2$ games the number of input neurons will be eight, and for $3 \times 3$ games the number of input neurons will be eighteen. The output layer must convey information about the decision made by the NN regarding the probability distribution over its own actions and therefore the number of neurons in this layer is equal to the size of the action set of each player i.e. two for $2 \times 2$ games and three for $3 \times 3$ games.

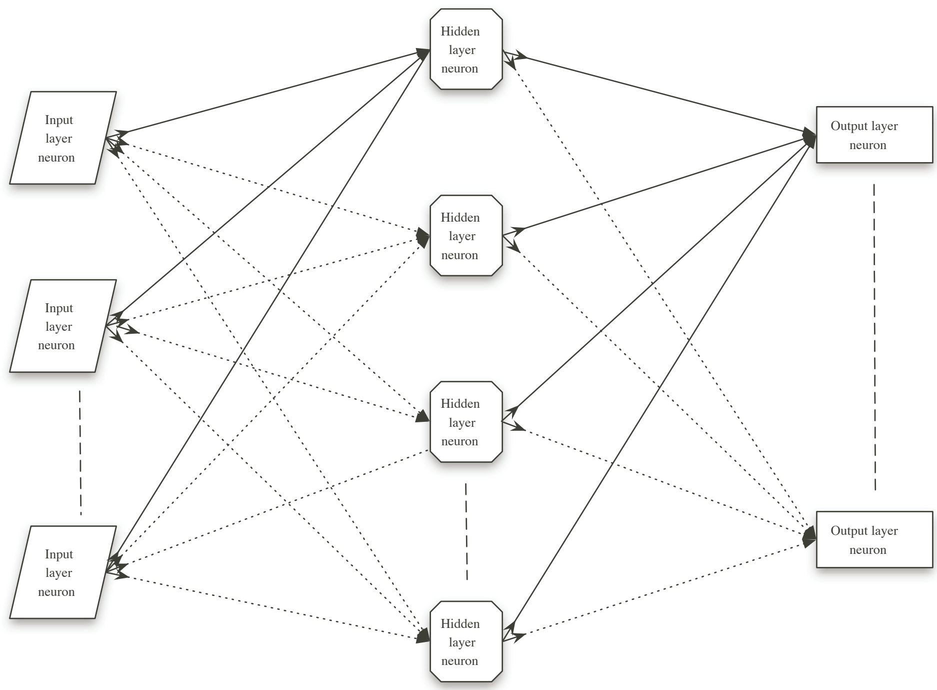
The number of hidden layers and the number of neurons in each layer can be chosen by the researcher according to different criteria, thereby altering the level of sophistication or bounded rationality of the NN[9]. For the rest of the analysis, the number of neurons, $v$, in each of $l$ hidden layers of a NN, will always be the same, therefore $\psi_i\{v,l\}$ denotes the complete topology of the $i$th NN agent[10].

The neurons in a standard feedforward NN are interconnected in the following way. Each neuron is connected to all the neurons in the previous layer (if such a layer exists) and all the neurons in the following layer (if such a layer exists), with neurons in the same layer not connected to each other. The input to a neuron is

[9] It is possible to simply connect the input layer to the output layer without any hidden layers intervening, resulting in a structure that is similar to a generalized linear model with the link function determined by the kind of non-linearity implemented in the output layer neurons.

[10] The number of neurons in the input and output layers are left out of this shorthand notation for simplicity as they can be inferred by the types of games being playing.

FIGURE 1.— Graphical representation of the structure and topology of feedforward neural networks




simply a weighted average of all the outputs of the neurons in the previous layer. These connection weights are what allow the NN to learn as they are constantly being adapted based on the performance of the NN.

As regards a NN's behavior in the game it is playing, information flows from the input layer to the hidden layers and finally to the output layer. However, as regards the learning mechanism of the NN and how it adapts its behavior, this process implicitly uses a backward flowing system, as exemplified by its name, the backpropagation rule. After making a decision the NN compares the values of the neurons at the output layer to the desired values and makes adjustments to all the connection weights in such a way that would reduce the error of the NN for that particular observation. The backpropagation learning rule uses the chain rule to assign the contribution of each neuron to the observed error, from which it is then possible to extract the necessary information regarding how to change each connection weight in order to reduce the overall error. Hence, each connection weight is changed according to a gradient descent method with the intent that the network successively approaches a state where the error function attains a global minimum.

The NN agents are assumed to consider the myopic best response to their opponent's action to be the desired response for each game. This approach is what is referred to as ex-post rational by Selten (1998). According to this principle, an economic agent will move or modify his action in the direction of ex-post best response to the temporally prior outcome. A popular example is that of an archer aiming at a target, whereby he observes whether his previous shot was to the left or right of the target and then adjusts his aim in the opposite direction. The principle of ex-post rationality replaces the Bayesian learning approach of ex-ante rationality, with a much simpler heuristic that is computationally more tractable as it does not require sophisticated dynamic optimization.

### 3.4. *Specifications of NN simulations*

In agent-based simulations, the dynamics and steady state properties can be affected by the values of model parameters. More specifically for this application, the number of agents in the population could affect the results of the simulations. This could be substantial when comparing a simulation with only two agents playing repeatedly with each other versus a simulation with more than two players. The existence of a population of only two players could facilitate the emergence of social conventions or focal points in coordination games compared to larger populations.

Another parameter that affects simulations is obviously the nature of the agents themselves and whether the population of NNs is homogeneous or heterogeneous. This research provides an obvious way of allowing for heterogeneous agents and modeling bounded rationality. Altering the network topology, specifically the number of layers and neurons per layer, affects the non-linear capabilities and sophistication of the neural networks. The non-linear capability of the agents is an increasing function of the number of neurons in each layer, holding the number of hidden layers constant, and also an increasing function of the number of hidden layers for a given number of neurons in each layer. The combined effect of simultaneously varying $v$ and $l$ in opposite directions will depend on the nature of the data and application as there is no steadfast rule relating the relative effects of such changes.

A common quandary researchers face in empirical applications of NNs is the selection of the network topology i.e. the number of layers and number of neurons to use. Walczak and Cerpa (1999) discuss the effects of variations in the topology and suggest heuristic principles in determining these parameters for practitioners. Increasing the number of layers leads to an increase in the closeness of fit as the NN is able to better deal with higher order or more complex problems. However there is a tradeoff for this better fit since this usually leads to worse generalization capability as the network output becomes less smooth and more prone to fitting noise. Increasing the number of neurons can lead to overfitting and loss of generalization whereas too small a number cannot provide enough flexibility for the network to learn and model a non-linear relationship.

In light of the above results and in particular the comparisons that must be made between different treatments, a number of different simulations must be examined. All the neural networks employed in the simulations will be fully connected feedforward networks incorporating hyperbolic tangent sigmoid transfer (or tansig) functions in the hidden layers and softmax neurons in the output layer. Readers not familiar with neural networks are referred to Appendix A.1 for a detailed discussion. Different simulations will be denoted by $\Psi$, where a subscript denotes whether the simulation is comprised of heterogeneous or homogeneous agents and a superscript denotes the size of the player's action space. All simulations with homogeneous agents, denoted by the subscript *hom*, employ NNs with three layers and 50 neurons in each layer i.e. $\psi\{50,3\}$. In all the heterogeneous agent simulations, denoted by the subscript *het*, the ten different NNs employed are listed below in order of decreasing bounded rationality or increasing sophistication:

$\psi_1\{\varnothing,\varnothing\}, \psi_2\{5,1\}, \psi_3\{5,2\}, \psi_4\{5,3\}, \psi_5\{20,1\}, \psi_6\{20,2\}, \psi_7\{20,3\}, \psi_8\{50,1\}, \psi_9\{50,2\}, \psi_{10}\{50,3\}$

The following list specifies all the simulations that will be discussed in the main body of this paper:

$$\begin{aligned}
\Psi^2_{hom} &= [\psi_1\{50,3\},...,\psi_{10}\{50,3\}] \\
\Psi^2_{het} &= [\psi_1\{\varnothing,\varnothing\}, \psi_2\{5,1\}, \psi_3\{5,2\}, \psi_4\{5,3\}, \psi_5\{20,1\}, \psi_6\{20,2\}, \psi_7\{20,3\}, \\
&\quad \psi_8\{50,1\}, \psi_9\{50,2\}, \psi_{10}\{50,3\}] \\
\Psi^3_{hom} &= [\psi_1\{50,3\},...,\psi_{10}\{50,3\}] \\
\Psi^3_{het} &= [\psi_1\{\varnothing,\varnothing\}, \psi_2\{5,1\}, \psi_3\{5,2\}, \psi_4\{5,3\}, \psi_5\{20,1\}, \psi_6\{20,2\}, \psi_7\{20,3\}, \\
&\quad \psi_8\{50,1\}, \psi_9\{50,2\}, \psi_{10}\{50,3\}]
\end{aligned}$$

These simulations can be segregated into three treatments: one treatment varies the size of the action space, another specifies whether NNs in a population are homogeneous or heterogeneous, and a third treatment, nested within the previous treatment, which varies the topology of the neural networks. The most simple agents simulated, $\psi\{\varnothing,\varnothing\}$, are essentially equivalent to a logit model as the lack of a hidden layer means that the inputs to the output layer neurons are simply a weighted average of payoffs which are then subjected to a softmax or logistic transformation. The most sophisticated NN included, $\psi\{50,3\}$, is the same as the ones used in the homogeneous treatment allowing for comparisons as to how the same sophisticated NN may behave differently when in the presence of a different population of opponents.

### 3.5. *Simulation details*

Each simulation documented in Section 3.4 will be implemented in Matlab (2007) for a total of one million generations, and for each generation all the NN agents in the population will be randomly paired with one another to play a randomly sampled game. Games were created by sampling the $2n^2$ payoffs of $n \times n$ games from a uniform distribution on [-1,1][11] thereby automatically ruling out games with weakly dominated ac-

[11] This normalization is the most conducive to NN learning when neurons use the tansig function as each input has the same influence, and inputs to the neurons have zero mean and are restricted to be on the support of the tansig function where the derivative is significantly different from zero. This helps avoid the problem of saturation which occurs when the output values of a neuron are close to 1 or -1 and leads to significantly slower learning. The reader is referred to LeCun et al. (1998) for a more detailed discussion of these issues.

TABLE I

GENERALIZED $3 \times 3$ GAME

| | | Player 2 | | |
|---|---|---|---|---|
| | | Left | Center | Right |
| Player 1 | Up | $x_{11}, x'_{11}$ | $x_{12}, x'_{12}$ | $x_{13}, x'_{13}$ |
| | Middle | $x_{21}, x'_{21}$ | $x_{22}, x'_{22}$ | $x_{23}, x'_{23}$ |
| | Down | $x_{31}, x'_{31}$ | $x_{32}, x'_{32}$ | $x_{33}, x'_{33}$ |

tions. The generalized form of such a $3 \times 3$ game is given in Table I.

Each network's parameters were randomly initialized according to the Nguyen and Widrow (1990) layer initialization function, which allows for faster learning and convergence. The standard back-propagation rule was used for adjustment of the NN weights, a detailed explanation of which can be found in Appendix A.2. After observing the opponent's choice for that game the NNs will assume that the desired output is the myopic best response.

The robustness of the simulations to different initial weights of the neural networks and different learning rates in the backpropagation algorithm is detailed in Appendix B. The conclusion of these analyses is that the results are very robust to changing these specifications and therefore it is sufficient to perform only one complete run of each simulation[12]. Also, Appendix E.1 examines the effects of changing the game payoff sampling scheme on NN behavior and finds that the results are qualitatively robust.

## 4. ANALYSIS AND RESULTS OF $2 \times 2$ GAMES

This section briefly highlights the most important results from performing simulations of NN agents in general $2 \times 2$ games, with further results and detailed analyses relegated to Appendix D.

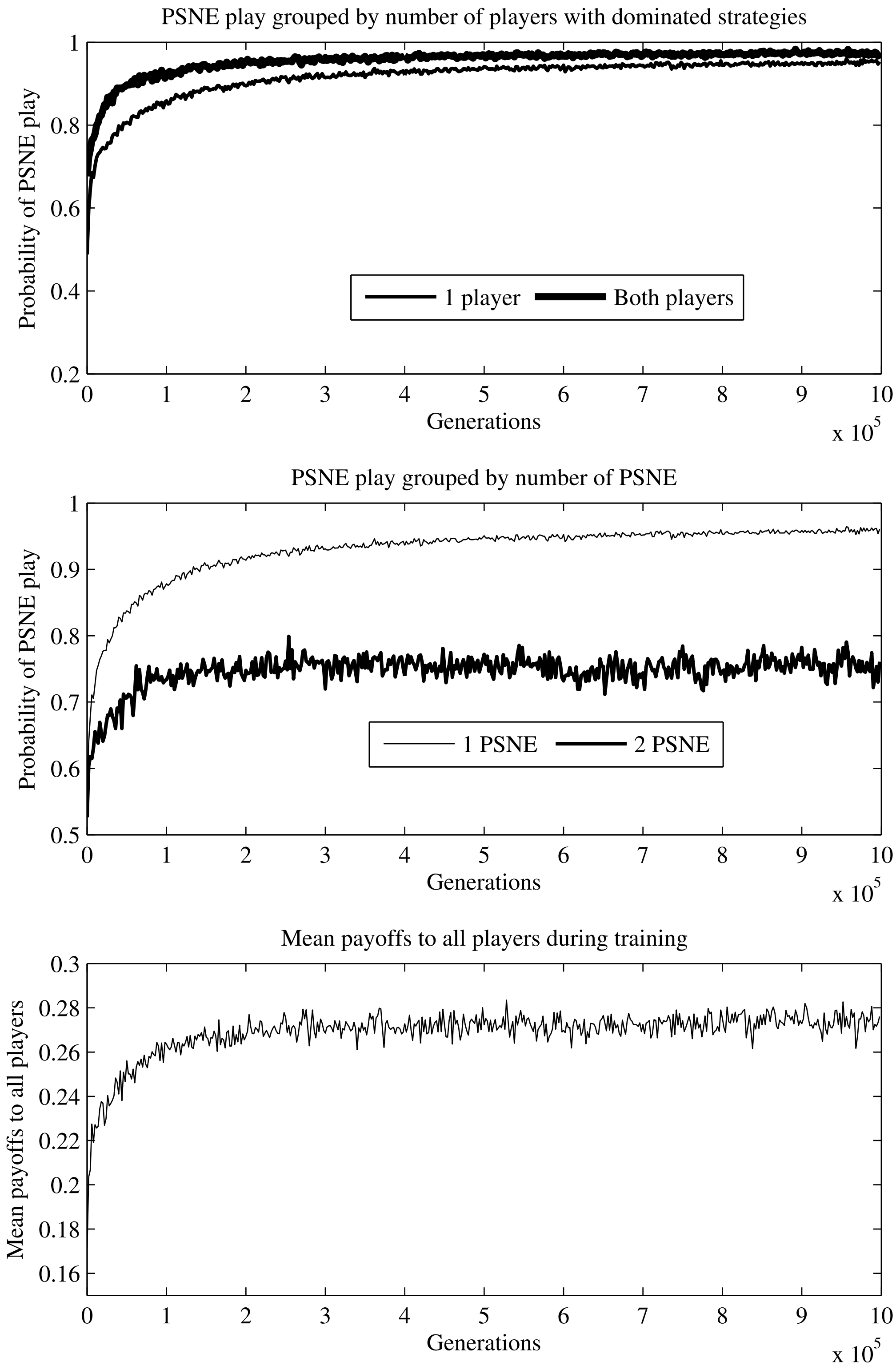
### 4.1. *Convergence to PSNE*

The top graph of Figure 2 plots the probability that a pair of networks will jointly play the PSNE of a game according to the number of players that have a dominant action. The middle graph segments the probability of PSNE behavior according to the number of PSNE of each game played.

In the cases of one or two players having a dominant action the networks are converging to playing a PSNE with near certainty. Note that in both of these cases there necessarily exists a unique PSNE for all such games. The case where no players have a dominant action is not included in this graph but these games must exhibit either zero PSNE or two PSNE, the latter is shown in the middle graph and the former requires specialized analysis that follows in Section 5.8 since it has only a MSNE.

[12]This is a fortunate result as the computational resources required to complete many runs of this simulation are forbidding given current technology.

FIGURE 2.— Behavior of the $\Psi^2_{hom}$ simulation during training



After one million generations the NNs perform exceptionally well in games with a unique PSNE, so that the Nash equilibrium prescription is jointly played by both networks almost 96% of the time[13], much higher than the random play baseline of 25%. Performance falls in games with two PSNE to roughly 75%, which is reasonable as the existence of two PSNE creates a coordination problem so that in many cases each network may go for a different equilibrium leading to a disequilibrium result. However, performance is still much higher than the random baseline prediction of 50%. As expected, convergence is faster for cases where both players have dominated strategies.

The question of convergence to a long-run stochastic steady state[14] can also be answered by observing the evolution of mean payoffs to players over generations as convergence would imply steady mean payoffs. The bottom graph in Figure 2 simply plots mean payoffs to all players versus the number of generations. The figure presents the mean payoffs of all ten players, averaged over a moving window of 10,000 generations to eliminate fluctuations and elucidate the underlying trend. The derivative of average payoffs over time with respect to the number of generations elapsed is positive but the second derivative is negative as the rate of learning appears to significantly slow down quite early in the training simulation. In particular, after the first 10,000 generations payoffs are already higher than 0.2, with the rate of increase tapering off and stabilizing, with the exception of random fluctuations which are to be expected due to the stochastic nature of the simulation. The plateau in payoffs is further evidence that the system has converged to or is in the close neighborhood of the stochastic steady state of this system.

### 4.2. *Equilibrium selection for* $2 \times 2$ *games in* $\Psi^2_{hom}$

To assist in the examination of equilibrium selection in $2 \times 2$ games with two PSNE a test data set of 1,000 generations was constructed. The main intention of this analysis is to ascertain whether risk[15] or payoff dominant equilibria are more likely to be played by the NNs. Table II compiles the probability of playing different types of PSNE. In particular, the probability of playing an equilibrium that is payoff dominant $p(PD)$, risk dominant $p(RD)$, payoff dominant but not risk dominant $p(PD\& \sim RD)$, risk dominant but not payoff dominant $p(RD\& \sim PD)$, both payoff and risk dominant $p(PD\&RD)$. Finally, the last column restricts analysis to games with distinct RD and PD equilibria, and calculates the ratio of the probability of playing the RD equilibrium to the probability of either the PD or RD equilibrium being played, abbreviated to $RD : PD + RD$.

Experimental research on the behavior of subjects playing coordination games finds that risk dominant

[13] Note, that these results refer to joint play of the PSNE, therefore assuming independence each player is choosing the unique PSNE action with a probability of roughly $\sqrt{0.96} = 0.98$.

[14] Due to the stochastic nature of the NNs' choices and the stochastic sampling of games there is no absorbing steady state.

[15] Readers interested in the method of computation of risk dominant equilibria are referred to Appendix C.2.

TABLE II
EQUILIBRIUM SELECTION BEHAVIOR OF NNS IN $2 \times 2$ GAMES

| $p(PD)$ | $p(PD)$ | $p(RD)$ | $p(PD\& \sim RD)$ | $p(RD\& \sim PD)$ |
|---|---|---|---|---|
| 0.5286 | 0.5286 | 0.6869 | 0.0826 | 0.6663 |
| $p(PD\&RD)$ | $p(PD\&RD)$ | $p(\sim PD\& \sim RD)$ | $RD : PD + RD$ | |
| 0.7259 | 0.7259 | 0.1055 | 0.8407 | |

equilibria are preferred. For example, Cooper et al. (1994) find that in a one-shot symmetric coordination game with two Pareto rankable pure strategy Nash equilibria, the Pareto inferior equilibrium is by far the most probable outcome. Repeated presentations with random rematching did not significantly affect these results as 97% of outcomes in the coordination games were still the Pareto inferior equilibrium. Cabrales et al. (2000) perform an experimental test on equilibrium selection in $2 \times 2$ games and they find that risk dominance is a good predictor of actual play, improving as the degree of asymmetry in payoffs between the two players increases. These results are corroborated by Straub (1995) who concludes that risk dominance is a better predictor of play than payoff dominance and that a necessary but not sufficient condition for coordination failure is the existence of a payoff dominated risk dominant equilibrium.

The results from the experimental literature are clearly replicated by the NNs which are converging to playing the risk dominant equilibrium, as the probability of playing the risk dominant equilibrium given that the other equilibrium is Pareto dominant, $RD : PD + RD$, is 0.8407. This is supported by the low value of $p(PD\& \sim RD)$, 0.0826, for selecting a payoff dominant equilibrium that it is not also risk dominant.

### 4.3. *Compliance with dominance principles*

The NNs have endogenously learned to play their dominant actions with a probability of 0.988, and to perform one-level iterated dominance, equivalent in this case to best responding to an opponent's dominant action, with a probability of 0.9736. This last result is particularly remarkable as it involves higher level reasoning or beliefs about an opponent's likely course of action.

## 5. ANALYSIS AND RESULTS OF $3 \times 3$ GAMES

### 5.1. *Homogeneity of trained NNs in* $\Psi^3_{hom}$

This section examines whether the NNs within a homogeneous agent simulation behave similarly or whether they have evolved to play differently, so that various types of players emerge in the population. Table III gives the Spearman cross-correlations of each NN's probabilistic response to a test set of 1,000 games. The average value is simply the average Spearman correlation coefficient for each possible pairing of two different NNs, the minimum and maximum values of the smallest and largest coefficients of any pair

TABLE III

SPEARMAN CROSS-CORRELATIONS OF INDIVIDUAL NN OUTPUT IN $\Psi^3_{hom}$

| | No. of PSNE | | | | |
|---|---|---|---|---|---|
| $\rho$ | Any | 0 | 1 | 2 | 3 |
| Average | 0.9352 | 0.8084 | 0.9409 | 0.9242 | 0.8982 |
| Minimum | 0.9188 | 0.738 | 0.9198 | 0.8892 | 0.7915 |
| Maximum | 0.9541 | 0.8928 | 0.9591 | 0.95 | 0.9814 |

are also presented. It is clear that the NNs behave quite similarly in general as the correlations are high and the minimum and maximum values quite close implying stability of the correlation coefficient. This result is evidence that there exists little if any systematic between-subjects heterogeneity since that would imply that each NN would play each game differently.

### 5.2. *Statistical behavior of NNs during training*

This section deals with the general dynamic behavior and characteristics of the neural networks' responses during the learning phase of two $3 \times 3$ game simulations, one with homogeneous agents and the other with heterogeneous agents. Figures 3 and 4 include three graphs showing different aspects of the networks' behavior. The top graph breaks down behavior according to the number of players with a dominant strategy and the middle graph plots behavior and its dependence on the number of PSNE each game exhibits. The lower graph plots mean payoffs to all players against the number of elapsed generations.

#### 5.2.1. *Behavior of $\Psi^3_{hom}$ during training*

From Figure 3 it is clear that the time series of the variables do not level off during the one million generations and still have significant positive first derivatives with respect to the number of elapsed generations. There is rapid learning initially whose rate of change over time tapers off leading to a roughly linear learning rate. The probability of joint PSNE play is quite similar for games with one, two or three PSNE, however this does not mean that the NNs exhibited the same amount of learning for each type of game because the appropriate baseline is the occurrence of joint PSNE play under random or uniformly distributed choice. According to random choice, for games with one, two and three PSNE the expectations of joint PSNE play are 0.11, 0.22 and 0.33 respectively implying that the networks found it easier to coordinate and play the PSNE the fewer the number of PSNE in the game. A conspicuous result from the graph is the variability around the underlying trend in the probability of PSNE play for games with three PSNE, which is much greater than the variability for games with fewer PSNE. Interestingly, periods of cooperation are often punctuated by short periods exhibiting a complete breakdown of cooperation so that the probability of playing one of the PSNE falls to the level expected from random behavior, 0.33.

FIGURE 3.— Behavior of the $\Psi^3_{hom}$ simulation during training

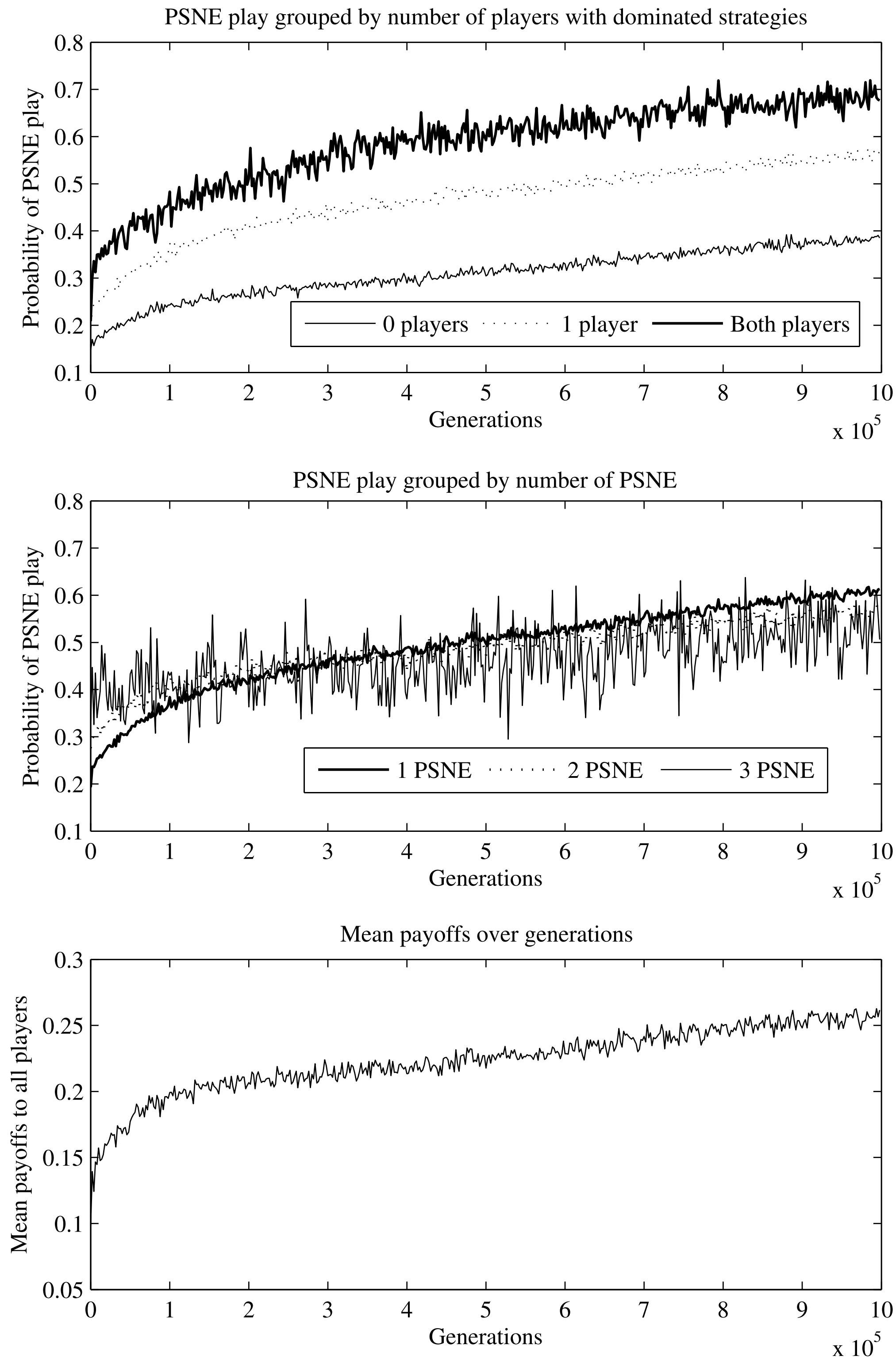
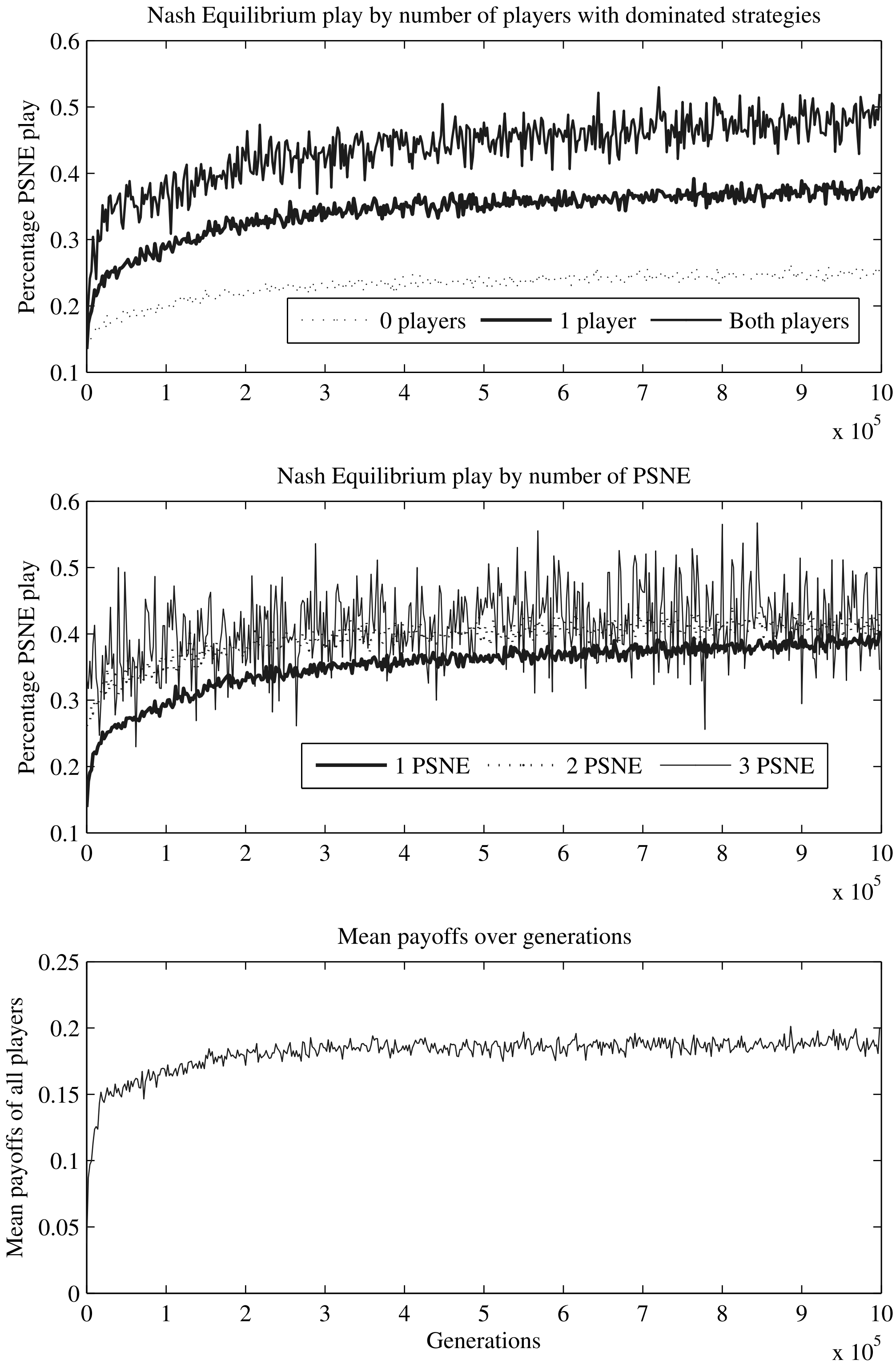


FIGURE 4.— Behavior of the $\Psi^3_{het}$ simulation during training



From the bottom figure it is clear that payoffs increase sharply the first few thousand generations and after 100,000 generations payoffs are increasing at a roughly linear rate over generations. The fact that payoffs are still increasing after one million generations verifies that the simulation has not yet reached a stochastic steady state[16].

#### 5.2.2. *Behavior of $\Psi^3_{het}$ during training*

The same variables as in the homogeneous case are plotted in Figure 4 and qualitatively the same results apply. The major difference however is that in all cases performance is lower in the heterogeneous simulation as the probability of PSNE play is lower in all cases, as are the average payoffs. This is a direct result of the lower level of average sophistication of the NNs in the heterogeneous population.

### 5.3. *Convergence to Pure Strategy Nash Equilibrium play*

#### 5.3.1. *Convergence by number of PSNE*

For the NNs facing $3 \times 3$ games performance drops compared to simulations of $2 \times 2$ games, the result of the increased computational complexities associated with the greater choice of possible actions. From Table IV, after one million generations, the networks jointly play a unique PSNE roughly 62% of the time which translates into a marginal probability of playing a PSNE action of roughly $\sqrt{0.62} = 0.787$. In contrast, Rey Biel (2004) found that in $3 \times 3$ games with a unique PSNE subjects played their PSNE action 79.625% of the time. Performance in games with two PSNE is considerably higher than the random baseline of 22.22%, whereas in games with three PSNE the networks jointly played one of the PSNE roughly 52% of the time, again appreciably higher than the random baseline of 33.33%. It is apparent that in moving to games with three actions instead of two, the complexity has increased appreciably so that the networks' behavior may be better described by heuristics rather than the NE prescription. More detailed analyses of the NNs' behavior in the case of multiple PSNE are given in Section 5.5.

The results for a heterogeneous agent simulation are provided in Table V. There is a clear positive relationship between the level of sophistication and the probability of joint PSNE play for all types of games. It should also be noted that comparing the results for $\psi_{10}\{50,3\}$ in the heterogeneous simulation to the results for the homogeneous simulation, where all agents are $\psi\{50,3\}$, the performance of this type of network is

[16] From the evidence at hand it seems that the simulations need to be run a significantly larger number of generations to achieve convergence. Current computational impediments to performing this simulation restricted the total number of generations that could be run both due to memory constraints and computational time required. It should be noted that evidence presented later in this paper will indicate that human behavior is approximated well with less than the one million generations performed in this analysis and therefore is adequate for the current purposes of this paper. Given the trends in increases in computational power and memory storage a simulation with enough generations to achieve convergence should be possible in the near future. Such a simulation would allow the examination of whether NNs can learn to converge to playing the PSNE in $3 \times 3$ games with arbitrarily small precision and will be performed by the author when it is computationally feasible.

TABLE IV

PROBABILITY OF JOINT NE PLAY IN $\Psi^3_{hom}$ ACCORDING TO NUMBER OF PSNE

| Number of PSNE | 1 | 2 | 3 |
|---|---|---|---|
| Probability of joint NE play | 0.6198 | 0.5762 | 0.5248 |

TABLE V

PROBABILITY OF JOINT NE PLAY IN $\Psi^3_{het}$ ACCORDING TO NUMBER OF PSNE

| | No. of PSNE | | |
|---|---|---|---|
| NN agents | 1 | 2 | 3 |
| $\psi_1\{\varnothing,\varnothing\}$ | 0.317 | 0.3552 | 0.3924 |
| $\psi_2\{5,1\}$ | 0.3447 | 0.3854 | 0.4217 |
| $\psi_3\{5,2\}$ | 0.354 | 0.4003 | 0.407 |
| $\psi_4\{5,3\}$ | 0.349 | 0.3825 | 0.4242 |
| $\psi_5\{20,1\}$ | 0.3907 | 0.4017 | 0.4176 |
| $\psi_6\{20,2\}$ | 0.4288 | 0.4323 | 0.4286 |
| $\psi_7\{20,3\}$ | 0.4251 | 0.4285 | 0.4458 |
| $\psi_8\{50,1\}$ | 0.4125 | 0.4026 | 0.4286 |
| $\psi_9\{50,2\}$ | 0.4379 | 0.4474 | 0.4458 |
| $\psi_{10}\{50,3\}$ | 0.4387 | 0.4579 | 0.4337 |

worse in the former simulation. This is not unreasonable as in the heterogeneous simulation $\psi\{50,3\}$ plays other networks of less sophistication who are not as adept at learning PSNE and therefore $\psi\{50,3\}$ has a smaller incentive to play and learn PSNE than in the homogeneous simulation.

5.3.2. *Convergence by number of players with dominated actions*

Investigating PSNE convergence by categorizing games according to the number of players that have a dominated action leads to the results in Tables VI and VII. The probability of observing joint PSNE play is strictly increasing in the number of players with dominant actions. This is intuitive because if both players have a dominant action then all that is required of them to achieve the PSNE is to play their dominant action. However, if only one player has a dominant action then in order for the PSNE to materialize one of the players must perform an iterated dominance calculation, which is clearly more complicated.

Examining the effects of sophistication in the heterogeneous simulation, it is clear from Table VII that the more sophisticated a NN the higher than probability of PSNE play for any given number of players with a dominant action. As before the performance of $\psi_{10}\{50,3\}$ is worse in the heterogeneous simulation compared to $\psi\{50,3\}$ in the homogeneous simulation.

TABLE VI

PROBABILITY OF JOINT PSNE PLAY IN $\Psi^3_{hom}$ ACCORDING TO NO. OF PLAYERS WITH DOMINANT ACTIONS

| No. of players with a dominant action | 0 | 1 | 2 |
|---|---|---|---|
| Probability of joint NE play | 0.4448 | 0.573 | 0.7044 |

TABLE VII

PROBABILITY OF JOINT PSNE PLAY IN $\Psi^3_{het}$ ACCORDING TO NO. OF PLAYERS WITH DOMINANT ACTIONS

| | No. of players with a dominant action | | |
|---|---|---|---|
| NN agents | 0 | 1 | 2 |
| $\psi_1\{\varnothing,\varnothing\}$ | 0.2403 | 0.301 | 0.4085 |
| $\psi_2\{5,1\}$ | 0.257 | 0.3471 | 0.4048 |
| $\psi_3\{5,2\}$ | 0.2685 | 0.3317 | 0.4305 |
| $\psi_4\{5,3\}$ | 0.2626 | 0.3358 | 0.4235 |
| $\psi_5\{20,1\}$ | 0.2854 | 0.3759 | 0.5 |
| $\psi_6\{20,2\}$ | 0.3115 | 0.4035 | 0.5254 |
| $\psi_7\{20,3\}$ | 0.3084 | 0.4144 | 0.5031 |
| $\psi_8\{50,1\}$ | 0.299 | 0.3874 | 0.519 |
| $\psi_9\{50,2\}$ | 0.3241 | 0.4145 | 0.5188 |
| $\psi_{10}\{50,3\}$ | 0.3179 | 0.4486 | 0.5393 |

### 5.4. *Dominance analyses*

This section examines how compliant the NNs are to dominance and iterated dominance principles and compares their behavior to experimental results. In particular, Stahl and Wilson (1995) performed an experiment where subjects played various $3 \times 3$ symmetric games with strictly and weakly dominated strategies. They found that only 4.86% of subjects' choices were strictly dominated and 5.42% of choices were either weakly or strictly dominated. Costa-Gomes et al. (2001) find that subjects played their dominant strategies almost 90% of the time in games with the following action structures: $2 \times 2$, $2 \times 3$ and $2 \times 4$. Of particular interest are their results on the relationship between NE play and the levels of iterated dominance required for a player to identify his/her equilibrium decision. As expected, compliance with NE behavior drops as the number of levels of iterated dominance increases, starting out at 89.2% for one level, falling to 65.5% for two levels and 15.3% for three levels of iterated dominance. Another study by Schotter et al. (1994) confirmed that subjects are much less likely to play a strongly dominated strategy than a weakly dominated strategy. Mookherjee and Sopher (1997) corroborate these results, finding that weakly dominated strategies are played only 2% to 8% of the time.

#### 5.4.1. *Analysis of* $\Psi^3_{hom}$ *simulation*

An analysis of the networks' behavior in games where dominance principles guide considerations of rationality follows. The computational requirements of adhering to own dominance are clearly less complex than those of iterated dominance. The former requires a player to determine only which of his own actions is dominant, whereas the latter makes the same requirement over an opponent's actions but then also requires a calculation of a best response.

Table VIII compiles the adherence to dominance principles for the homogeneous simulation. Given that

TABLE VIII

PROBABILITY OF NN COMPLIANCE IN $\Psi^3_{hom}$ WITH DOMINANCE AND ITERATED DOMINANCE PRINCIPLES

| | Dominance | It. dominance |
|---|---|---|
| Probability of compliance | 0.9256 | 0.8209 |

TABLE IX

PROBABILITY OF NN COMPLIANCE WITH DOMINANCE AND ITERATED DOMINANCE PRINCIPLES IN $\Psi^3_{het}$

| NN agent | Probability dominant play | Probability best responding to opponent |
|---|---|---|
| $\psi_1\{\varnothing,\varnothing\}$ | 0.8804 | 0.4637 |
| $\psi_2\{5,1\}$ | 0.8828 | 0.521 |
| $\psi_3\{5,2\}$ | 0.8893 | 0.5311 |
| $\psi_4\{5,3\}$ | 0.8722 | 0.5315 |
| $\psi_5\{20,1\}$ | 0.9168 | 0.6504 |
| $\psi_6\{20,2\}$ | 0.9317 | 0.7198 |
| $\psi_7\{20,3\}$ | 0.9422 | 0.7356 |
| $\psi_8\{50,1\}$ | 0.9271 | 0.6828 |
| $\psi_9\{50,2\}$ | 0.9413 | 0.7513 |
| $\psi_{10}\{50,3\}$ | 0.9574 | 0.7719 |

the NNs were not predisposed or preprogrammed in any way to follow dominance principles the results are impressive. The probability of complying with dominance, 0.9256 is extremely high. Although iterated dominance is violated more often than simple dominance, the NNs are still quite competent, adhering to this principle with a probability of 0.8209.

#### 5.4.2. *Analysis of $\Psi^3_{het}$ simulation*

The most reasonable hypothesis regarding the dominance behavior of the bounded rational networks is that the NNs are more likely to comply with dominance and iterated dominance theoretical prescriptions the more sophisticated they are. The rationale is that these prescriptions are essentially a discontinuous function of payoffs, and discontinuities in general require a large number of continuous functions, hence neurons, to be approximated well. This hypothesis is upheld by the results given in Table IX where sophistication clearly increases the probability of adhering to dominance and iterated dominance principles. The probability of playing a dominant action increases from a minimum value of roughly 0.88 for $\psi_1\{\varnothing,\varnothing\}$ to almost 0.96 for the most sophisticated network, $\psi_{10}\{50,3\}$, and even more impressive is the respective increase in the probability of complying with iterated dominance which rises from 0.4637 to 0.7719.

### 5.5. *Equilibrium selection*

In $3\times3$ games the possible multiplicity of PSNE leads to a question of equilibrium selection in games with two or three PSNE. Equilibrium selection is an important issue in the field and has inspired a number of theoretical refinements to NE and experimental studies. In Cooper et al. (1990) a number of hypotheses are

TABLE X

EQUILIBRIUM SELECTION BEHAVIOR OF NNS IN $\Psi^3_{hom}$

| | Probability of playing types of equilibria | | | |
|---|---|---|---|---|
| No. of PSNE | $p(PD)$ | $p(RD)$ | $p(PD\& \sim RD)$ | $p(RD\& \sim PD)$ |
| 2 | 0.3527 | 0.4236 | 0.1809 | 0.4111 |
| 3 | 0.2282 | 0.295 | 0.1389 | 0.2841 |

| | Probability of playing types of equilibria | | |
|---|---|---|---|
| No. of PSNE | $p(PD\&RD)$ | $p(\sim PD\& \sim RD)$ | $RD : PD + RD$ |
| 2 | 0.4511 | 0.0696 | 0.6577 |
| 3 | 0.3649 | 0.0868 | 0.6428 |

tested regarding play in $3 \times 3$ coordination games with Pareto rankable equilibria. They discovered that the hypothesis that play will generally fall into one of the Nash equilibria of the game was valid. However, the hypothesis that play will converge to the Pareto dominant equilibrium was refuted as subjects often locked into the Pareto inferior equilibrium. Even more interesting is the result that the Nash equilibrium that was played could be influenced by the magnitude of payoffs resulting from an opponent's strictly dominated action, implying that players at some point must have been placing a positive probability on the play of a dominated action by their opponents. Haruvy and Stahl (2004) find that for a set of $3 \times 3$ games only 8.4% of the time did the payoff dominant equilibrium predict actual experimental behavior, whereas the prediction accuracy of the risk dominant equilibrium was 62.4%.

Two test sets of 1,000 generations each were created which differed only as to the number of PSNE of the games included: one test set consisted only of games with two PSNE and the other only of games with three PSNE. The trained neural networks were then presented with these test sets and their behavior was documented with the intent to examine whether the experimental evidence in favor of risk dominant equilibria is upheld.

### 5.5.1. *Equilibrium selection for $3 \times 3$ games in $\Psi^3_{hom}$*

Table X provides the probability of playing different types of equilibria[17]. In games with two PSNE, a RD equilibrium is roughly 1.2 times more likely to be achieved than a PD equilibrium. The most likely equilibrium to be achieved, with a probability of 0.4511, is one that is both payoff and risk dominant. The $RD : PD + RD$ value of 0.6577 supports the conclusion that risk dominance is more important than payoff dominance in equilibrium selection in games with two PSNE.

Turning to games with three PSNE the results are not qualitatively different from games with two PSNE.

[17] Computing the payoff dominant equilibria is trivial for $2 \times 2$ and $3 \times 3$ games, as is the calculation of risk dominant equilibria in $2 \times 2$ games. However, the calculation of risk dominant equilibria for $3 \times 3$ games is more involved and the required computations are specified in Appendix C.2.

Again the most likely equilibrium is one that is both payoff and risk dominant, but if risk dominance and payoff dominance elect different PSNE, risk dominance is still significantly more likely. An overall trend is that the probabilities of playing all types of equilibria are significantly less in games with three PSNE compared to two PSNE. The obvious explanation of this is that the addition of another PSNE makes it even more difficult for players to coordinate as there are now three instead of only two possible actions which support an equilibrium.

In conclusion, the results are in line with the experimental findings of risk dominant equilibria being more likely than payoff dominant equilibria, although quantitatively the preference is not as strong as in the experimental studies.

#### 5.5.2. *Equilibrium selection for* $3 \times 3$ *games in* $\Psi^3_{het}$

The statistics in Table XI are the probabilities of attaining the equilibria specified in the columns given that one of the players of the game is the one specified in the rows. For example, the intersection of the first row and first column gives the probability of jointly attaining a PD equilibrium when one of the two players is $\psi_1\{\varnothing,\varnothing\}$. It is apparent that the level of sophistication of a bounded rational agent impacts equilibrium selection in games with two PSNE moderately, and only minimally in games with three PSNE. In games with two PSNE, as sophistication increases the probability of playing a risk dominant equilibrium increases in general, as documented by the increases in $p(RD)$, $p(RD\& \sim PD)$ and the increase in $RD : PD + RD$.

### 5.6. *Payoffs analysis*

This section assesses the variations in payoff performance of NN agents trained in homogeneous and heterogeneous simulations of $3 \times 3$ games.

#### 5.6.1. *Analysis of individual NNs in the* $\Psi^3_{hom}$ *simulation*

The average payoffs to the NNs are given in Table XII. The highest average payoffs occurred for games with a single PSNE, followed by games with two and three PSNE and finally games with a unique MSNE.

#### 5.6.2. *Analysis of individual NNs in the* $\Psi^3_{het}$ *simulation*

A hypothesis that more sophisticated networks would be capable of achieving higher payoffs is supported in simulations of a test set of $3 \times 3$ games, as communicated by Table XIII. The lowest average payoff accomplished is 0.1459 by $\psi_1\{\varnothing,\varnothing\}$ whereas the highest payoffs, roughly 0.21, are attributed to the networks with 50 neurons in each layer, $\psi_8$, $\psi_9$ and $\psi_{10}$. There does not appear to be any significant variation in payoffs across networks for games with zero PSNE, contrasting sharply with the large variation in payoffs for games

TABLE XI.— Individual NN equilibrium selection behavior in $\Psi^3_{het}$

| | | Probability of playing types of equilibria | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | NN agent | $p(PD)$ | $p(RD)$ | $p(PD\& \sim RD)$ | $p(RD\& \sim PD)$ | $p(PD\&RD)$ | $p(\sim PD\& \sim RD)$ | $RD : PD+RD$ |
| Games with 2 PSNE | $\psi_1\{\varnothing,\varnothing\}$ | 0.2223 | 0.2477 | 0.1472 | 0.2386 | 0.2666 | 0.0928 | 0.6022 |
| | $\psi_2\{5,1\}$ | 0.2322 | 0.2608 | 0.1437 | 0.2519 | 0.2811 | 0.0913 | 0.6028 |
| | $\psi_3\{5,2\}$ | 0.2416 | 0.2807 | 0.1533 | 0.2716 | 0.2978 | 0.0884 | 0.5966 |
| | $\psi_4\{5,3\}$ | 0.2319 | 0.253 | 0.1567 | 0.2426 | 0.2763 | 0.0919 | 0.5674 |
| | $\psi_5\{20,1\}$ | 0.2615 | 0.296 | 0.1567 | 0.2873 | 0.316 | 0.0865 | 0.6314 |
| | $\psi_6\{20,2\}$ | 0.2448 | 0.2924 | 0.1449 | 0.2874 | 0.3013 | 0.0871 | 0.6261 |
| | $\psi_7\{20,3\}$ | 0.2442 | 0.3072 | 0.1366 | 0.3057 | 0.3115 | 0.0855 | 0.6732 |
| | $\psi_8\{50,1\}$ | 0.2652 | 0.3016 | 0.1394 | 0.2867 | 0.3337 | 0.0863 | 0.6538 |
| | $\psi_9\{50,2\}$ | 0.2575 | 0.3007 | 0.1531 | 0.2964 | 0.3109 | 0.086 | 0.6465 |
| | $\psi_{10}\{50,3\}$ | 0.2493 | 0.3003 | 0.1592 | 0.2972 | 0.307 | 0.0858 | 0.6241 |
| Games with 3 PSNE | $\psi_1\{\varnothing,\varnothing\}$ | 0.1585 | 0.2072 | 0.1021 | 0.201 | 0.251 | 0.099 | 0.6472 |
| | $\psi_2\{5,1\}$ | 0.1703 | 0.2204 | 0.1086 | 0.2134 | 0.2645 | 0.0971 | 0.6265 |
| | $\psi_3\{5,2\}$ | 0.1631 | 0.2213 | 0.0989 | 0.2146 | 0.275 | 0.0972 | 0.6545 |
| | $\psi_4\{5,3\}$ | 0.1691 | 0.2192 | 0.1115 | 0.2127 | 0.2643 | 0.0972 | 0.6038 |
| | $\psi_5\{20,1\}$ | 0.1659 | 0.2384 | 0.1048 | 0.2334 | 0.2734 | 0.0949 | 0.6605 |
| | $\psi_6\{20,2\}$ | 0.1796 | 0.2261 | 0.104 | 0.2165 | 0.2835 | 0.0965 | 0.643 |
| | $\psi_7\{20,3\}$ | 0.176 | 0.2338 | 0.1091 | 0.2301 | 0.2552 | 0.0955 | 0.6335 |
| | $\psi_8\{50,1\}$ | 0.1717 | 0.2263 | 0.1029 | 0.2213 | 0.2544 | 0.0965 | 0.6553 |
| | $\psi_9\{50,2\}$ | 0.1767 | 0.2373 | 0.1128 | 0.2309 | 0.2785 | 0.0948 | 0.625 |
| | $\psi_{10}\{50,3\}$ | 0.1593 | 0.2285 | 0.1067 | 0.2249 | 0.2637 | 0.0962 | 0.6522 |

TABLE XII

COMPARISON OF MEAN PAYOFFS OF NNS IN $\Psi^3_{hom}$ GROUPED BY NO. OF PSNE

| | No. of PSNE | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| Payoffs | 0.0928 | 0.3291 | 0.2429 | 0.1522 |

TABLE XIII

INDIVIDUAL NN MEAN PAYOFFS IN $\Psi^3_{het}$

| | No. of PSNE | | | | |
|---|---|---|---|---|---|
| NN agents | 0 | 1 | 2 | 3 | Any |
| $\psi_1\{\varnothing,\varnothing\}$ | 0.1077 | 0.1681 | 0.1239 | 0.0971 | 0.1459 |
| $\psi_2\{5,1\}$ | 0.1004 | 0.1904 | 0.1471 | 0.103 | 0.1618 |
| $\psi_3\{5,2\}$ | 0.0882 | 0.1881 | 0.1514 | 0.0976 | 0.1586 |
| $\psi_4\{5,3\}$ | 0.0849 | 0.1988 | 0.1391 | 0.1072 | 0.1618 |
| $\psi_5\{20,1\}$ | 0.1169 | 0.2371 | 0.1761 | 0.11 | 0.1981 |
| $\psi_6\{20,2\}$ | 0.1045 | 0.2447 | 0.1663 | 0.1097 | 0.1973 |
| $\psi_7\{20,3\}$ | 0.1394 | 0.254 | 0.1577 | 0.0949 | 0.2098 |
| $\psi_8\{50,1\}$ | 0.1273 | 0.2643 | 0.1599 | 0.1048 | 0.2136 |
| $\psi_9\{50,2\}$ | 0.0985 | 0.2616 | 0.1724 | 0.107 | 0.2081 |
| $\psi_{10}\{50,3\}$ | 0.112 | 0.2526 | 0.1752 | 0.1159 | 0.2059 |

with a single PSNE. The rate of increase in payoffs with sophistication is smaller for games with two PSNE, compared to games with a unique PSNE. The payoffs for games with three PSNE do not vary systematically with the level of sophistication, with the simpler networks achieving roughly the same results as much more sophisticated NNs[18].

Finally, an interesting comparison can be made between the payoffs of $\psi_{10}\{50,3\}$ and the payoffs of networks of the same topology in the homogeneous population of NNs. The conclusion is that a sophisticated NN earns higher payoffs for games that have at least one PSNE, when playing against a population of other sophisticated NNs compared to playing against a population of less sophisticated agents. Hence, a sophisticated agent if given the choice would prefer to play against other agents of similar sophistication, instead of a population of less sophisticated agents.

This analysis, however does not allow for detailed examination of payoff variations according to the sophistication of both opponents in a game. The proposed solution is to perform spatial regression on the payoff data by implementing an ordinal distance measure based on the relative sophistication of the NNs. It is assumed that sophistication rankings follow the numbering $i$ in the shorthand symbol for NNs $\psi_i$, so that the sophistication of $\psi_i$ is $S_i = i$. The payoffs from all possible pairings of NN agents, are used as the observations of the dependent variable in equation 1 where $\pi_i^j$ is the average payoff to NN $i$ whenever it encountered NN

[18]The payoffs for games with three PSNE were calculated as the average of the last 500,000 generations of the training schedule because of the large variability in behavior in these games as was pointed out in Section 5.2, rendering the test set results problematic in drawing conclusions about relative performance.

$j$.

$$\text{(1)} \qquad \pi_i^j = \alpha + \beta_{as}((S_i + S_j)/2) + \beta_{rs}(S_i - S_j) + \beta_{abs}(abs(S_i - S_j)) + \varepsilon_i^j$$

Three independent variables are used based on the levels of sophistication of the NNs in each game matching. The variable $(S_i + S_j)/2$ is the average sophistication of the pair of NNs, with an estimated coefficient denoted by $\beta_{as}$. The variable $S_i - S_j$, is simply a measure of the difference in sophistication, or relative sophistication, between the two agents and its coefficient is $\beta_{rs}$. Finally, the variable $abs(S_i - S_j)$ is an absolute measure of the similarity of the two neural networks and its associated coefficient is $\beta_{abs}$.

This model allows the testing of three hypotheses. The first hypothesis is that the higher the average intelligence of two opponents, the higher payoffs are i.e. $\beta_{as} > 0$. The second hypothesis, that the more sophisticated a network is relative to its opponent the higher its payoffs, implies that $\hat{\beta}_{rs} > 0$. The third hypothesis is that payoffs are higher the more similar two networks are in terms of sophistication, implying $\beta_{abs} < 0$. Such a hypothesis could be reasonable in games where there exist more than one PSNE and coordinating becomes an important issue. It may be easier for two agents to coordinate if they are of a comparable level of sophistication since their behavior will likely be similar. This model is run separately on each subset of the games, as defined by the number of PSNE, and the results are exhibited in Table XIV.

The data used for games with zero, one and two PSNE was a test set during which there was no NN learning, but because of the high variability of behavior in games with three PSNE, the last 500,000 generations of the training set were used instead to estimate this equation. For games with zero, one and two PSNE, all estimated coefficients have the same sign, $\beta_{as} > 0$, $\beta_{rs} > 0$ and $\beta_{abs} < 0$ but the latter is only significantly less than zero in games with zero PSNE. For games with three PSNE increasing average sophistication still has a statistically significant positive effect on payoffs, but now $\beta_{rs} < 0$ and $\beta_{abs} > 0$. Hence, the more dissimilar opponents are the higher the payoffs, and the higher the relative sophistication of an agent the lower payoffs are.

Examining the incentives a player has to behave more sophisticated (if possible) it is easy to show that for all games with less than three PSNE, $d\pi_i^j/dS_i > 0$ i.e. players have an incentive to become even more sophisticated, assuming their opponent's behavior does not change. Note, that this conclusion holds regardless of whether the player under examination is initially less or more sophisticated than his opponent. Turning to whether agents playing games with less than three PSNE would prefer to play against sophisticated opponents it can be shown that if $S_i > S_j$ then $d\pi_i^j/dS_j > 0$ but if $S_i < S_j$ then $d\pi_i^j/dS_j < 0$, which means that players benefit if the sophistication of their opponent increases only as long as they are still more sophisticated than

TABLE XIV

DEPENDENCE OF PAYOFFS ON THE DIFFERENCE IN SOPHISTICATION OF NN OPPONENTS IN $\Psi^3_{het}$

| # of PSNE | Variable | Coef. | Bias | Boostrap s.e. | $lower_{95\%}$ | $upper_{95\%}$ |
|---|---|---|---|---|---|---|
| 0 | $(S_i+S_j)/2$ | 0.0106 | 0.0000 | 0.0019 | 0.0070 | 0.0145 |
| | $S_i-S_j$ | 0.0030 | 0.0000 | 0.0008 | 0.0013 | 0.0046 |
| | $abs(S_i-S_j)$ | -0.0044 | 0.0000 | 0.0016 | -0.0077 | -0.0012 |
| | $constant$ | 0.1416 | -0.0001 | 0.0111 | 0.1221 | 0.1664 |
| | $Wald\,\chi^2(2)$ | 44.0000 | | | | |
| | $p(\chi^2(2))>0$ | 0.0000 | | | | |
| | $R^2$ | 0.3426 | | | | |
| 1 | $(S_i+S_j)/2$ | 0.0080 | 0.0000 | 0.0012 | 0.0057 | 0.0106 |
| | $S_i-S_j$ | 0.0041 | 0.0000 | 0.0006 | 0.0030 | 0.0053 |
| | $abs(S_i-S_j)$ | -0.0006 | 0.0000 | 0.0011 | -0.0027 | 0.0016 |
| | $constant$ | 0.1484 | 0.0002 | 0.0089 | 0.1303 | 0.1647 |
| | $Wald\,\chi^2(2)$ | 103.5000 | | | | |
| | $p(\chi^2(2))>0$ | 0.0000 | | | | |
| | $R^2$ | 0.5218 | | | | |
| 2 | $(S_i+S_j)/2$ | 0.0061 | 0.0001 | 0.0027 | 0.0007 | 0.0115 |
| | $S_i-S_j$ | 0.0034 | 0.0000 | 0.0010 | 0.0016 | 0.0058 |
| | $abs(S_i-S_j)$ | -0.0018 | 0.0001 | 0.0022 | -0.0059 | 0.0029 |
| | $constant$ | 0.1479 | -0.0009 | 0.0161 | 0.1156 | 0.1782 |
| | $Wald\,\chi^2(2)$ | 18.1300 | | | | |
| | $p(\chi^2(2))>0$ | 0.0004 | | | | |
| | $R^2$ | 0.1820 | | | | |
| 3 | $(S_i+S_j)/2$ | 0.0043 | 0.0000 | 0.0011 | 0.0024 | 0.0065 |
| | $S_i-S_j$ | -0.0011 | 0.0000 | 0.0005 | -0.0023 | -0.0002 |
| | $abs(S_i-S_j)$ | 0.0022 | 0.0000 | 0.0011 | 0.0001 | 0.0043 |
| | $constant$ | 0.0775 | 0.0000 | 0.0076 | 0.0626 | 0.0917 |
| | $Wald\,\chi^2(2)$ | 29.27 | | | | |
| | $p(\chi^2(2))>0$ | 0 | | | | |
| | $R^2$ | 0.197 | | | | |

their opponent.

In games with three PSNE, $d\pi_i^j/dS_i > 0$ if $S_i > S_j$ but $d\pi_i^j/dS_i < 0$ if $S_i < S_j$, whereas $d\pi_i^j/dS_j > 0$ for all values of $S_i$ and $S_j$ implying that agents always benefit when their opponent increases in sophistication.

### 5.7. *Rule extraction and heuristics*

Many cite the main drawback of NNs to be the fact that they are black boxes, in the sense that one cannot be sure exactly what calculations are going on inside. This is a direct consequence of the fact that NNs encode information in a distributed structure making it hard to extract any simple, comprehensible understanding of how NNs come to the conclusions they do. As a response to this, researchers have studied how simple rules can be extracted from a NN[19]. The breadth of research for NNs performing regression or function approximation is more limited than for networks that are performing classification problems. This is partly due to the fact that it is harder to come up with a simple set of rules for a regression problem due to the continuity of the networks' output in contrast to classification problems where the output is usually a small set of classes.

There is an inherent trade off when extracting rules from NNs between accuracy and comprehensibility. Accuracy is a measure of how good the fit of the extracted rules is compared to the actual NN output. However, comprehensibility dictates that rules be as simple and as few as possible which necessarily leads to a detrimental effect on the fit of the extracted rules. The actual form of the model will be dictated more by our demand that it be a simple and easily interpretable model which can examine whether specific salient strategies have been discovered by the NN. Hence, in this particular application it is prudent to place more weight on comprehensibility, given that a satisfactory level of accuracy is ensured.

A direct comparison between NN and human behavior can be performed by implementing the heuristics that subjects have been found to use in experimental research. Also, since in this paper our interest lies in whether the NN has learned game strategies such as avoiding to play dominated strategies and performing basic iterated dominance, a regression model will be built incorporating these elements as well as the standard heuristics in the experimental literature.

Many experimental studies provide statistical evidence of the use of particular heuristics specified by the researcher. Hence, for the sake of comparison it is desirable to decompose the NN behavior in terms of the heuristics commonly employed in these studies. Stahl and Wilson (1995) proposed to model the heterogeneity of players by limiting their degree of sophistication to three levels. They defined Level-0 players as agents whose play was a simple uniform distribution over pure actions. Higher level players were defined as best responding to the assumption that their opponent was one level below them in sophistication e.g. a Level-1

[19] An excellent survey and taxonomy of rule extraction methods is presented in Huysmans et al. (2006).

player best responded to a Level-0 player. Subsequent studies such as Costa-Gomes et al. (2001) also used similar types and added some other types to their analysis. The relationship between the heuristics employed in this paper and these previous papers is clarified below:

The heuristics are the following (henceforth referred to as the heuristic variables):

**Minimax** This heuristic, first expounded by von Neumann, states that players will try to minimize the maximum loss that an opponent can force them to suffer. This heuristic is the same as *pessimistic* or *maximin* in Costa-Gomes et al. (2001).

**Best response to minimax** This heuristic best responds to the assumption that an opponent is playing according to minimax. This heuristic was not included in previous studies to the best of the author's knowledge.

**Maximax** A player following this heuristic will play the action which includes the highest possible own payoff. This heuristic is also referred to as *optimistic* in Costa-Gomes et al. (2001).

**Best response to maximax** This heuristic best responds to an opponent following the maximax heuristic. This heuristic was not included in previous studies to the best of the author's knowledge.

***L*1** This heuristic states that a player will best respond to a uniform prior distribution over his opponent's actions. Also referred to as *naive* in Costa-Gomes et al. (2001).

***L*2** This heuristic best responds to an opponent that plays according to the $L1$ heuristic. The same heuristic exists in Costa-Gomes et al. (2001).

***L*3** This heuristic best responds to an opponent that plays according to the $L2$ heuristic. This heuristic is an extension to the types in Stahl and Wilson (1995) and Costa-Gomes et al. (2001), where analysis is limited to Level-2 types.

***L*4** This heuristic best responds to an opponent that plays according to the $L3$ heuristic. This heuristic is another extension of the types in Stahl and Wilson (1995) and Costa-Gomes et al. (2001).

On the basis of widespread experimental evidence of conformance to dominance and iterated dominance this paper will include dominance variables concurrently with these heuristic variables in models approximating NN behavior.

Regarding the classification of experimental subjects according to the type of heuristic that best describes their behavior, Stahl and Wilson (1995) estimated the frequencies of types using a mixture model and found that 17.5% were $L0$ types, 20.7% were $L1$, 2.1% were $L2$ and 43.1% were wordly Nash types, who assume that all other players are either $L0$, $L1$, $L2$ or Nash equilibrium players and best responds to prior beliefs about the frequency of each of these types[20]. The importance of $Ln$ heuristics in explaining behavior is also corroborated by the seminal study of Costa-Gomes et al. (2001) which also collected data on subjects'

[20] This type of player will therefore behave quite often as $L1$, $L2$, $L3$ as these are the best responses to $L0$, $L1$, $L2$.

TABLE XV

LOGIT REGRESSION OF NN OUTPUT IN THE $\Psi^3_{hom}$ SIMULATION

| | Coef. | Std. Err. | $t$-stat. | $p$-value | $lower_{95\%}$ | $upper_{95\%}$ |
|---|---|---|---|---|---|---|
| $\beta_{minimax}$ | -0.005 | 0.009 | -0.510 | 0.609 | -0.022 | 0.013 |
| $\beta_{br-minimax}$ | 0.071 | 0.008 | 9.170 | 0.000 | 0.056 | 0.087 |
| $\beta_{maximax}$ | 0.073 | 0.008 | 8.950 | 0.000 | 0.057 | 0.089 |
| $\beta_{br-maximax}$ | 0.084 | 0.008 | 10.860 | 0.000 | 0.069 | 0.100 |
| $\beta_{L1}$ | 0.217 | 0.009 | 23.320 | 0.000 | 0.199 | 0.235 |
| $\beta_{L2}$ | 0.460 | 0.008 | 55.390 | 0.000 | 0.443 | 0.476 |
| $\beta_{L3}$ | 0.322 | 0.009 | 36.780 | 0.000 | 0.305 | 0.340 |
| $\beta_{L4}$ | 0.260 | 0.009 | 29.640 | 0.000 | 0.243 | 0.278 |
| $\beta_{it.dom.}$ | 0.244 | 0.023 | 10.650 | 0.000 | 0.199 | 0.289 |
| $\beta_{dominated}$ | -0.512 | 0.017 | -30.510 | 0.000 | -0.545 | -0.479 |
| $LR\,\chi^2(8)$ | 39797.56 | | | | | |
| $Pseudo-R^2$ | 0.1811 | | | | | |

attention to payoff information, thereby allowing for greater power and precision in classifying subjects into types based on their information search patterns. Using both decision and search information, the two most important heuristics were by far $L1$ and $L2$, whose frequencies of occurrence in the subject pool were 44.8% and 44.1% respectively.

### 5.7.1. *Learned heuristics of networks in the $\Psi^3_{hom}$ simulation*

Conditional logit regressions on the whole dataset incorporating all of these heuristics are exhibited in Table XV. Also included as independent variables are dummies that denote whether an action is dominated by at least one other action in a player's action set, denoted by $dominated$. The variable, $it.dom.$, captures the possibility of the networks performing one-level of iterated dominance in games that are solvable by such a strategy i.e. for $3 \times 3$ games this requires the existence of a dominant action, not just a dominated action. Alternatively, it is defined as playing the best response to an opponent's dominant strategy, which guarantees the existence of a definitive unique best response. These independent variables (henceforth referred to as the dominance variables) do not make prescriptions in all types of $3 \times 3$ games for the following reason. Since a dominated action cannot support a Nash equilibrium the variable $dominated$ is not defined in games with three PSNE. For the same reason, the $it.dom.$ variable can only support a single PSNE.

The heuristics $L3$ and $L4$ have been omitted from the analysis of games with three PSNE, because the solutions to $L1$ and $L3$ are necessarily the same in this case, as are the solutions to $L2$ and $L4$. The intuitive reasoning for this result is the following. Let the set of games with three PSNE be partitioned into two sets based on the $L1$ responses of both players. In the first case, assume that the $L1$ solutions of both players lead to a PSNE implying that they are best responses to each other. By definition the $L2$ solution for each player will be the best response to the opponent's $L1$ solution. But a best response to a PSNE is the PSNE action itself

TABLE XVI

*Ln* HEURISTICS IN GAMES WITH THREE PSNE

| | | $L1$, $L3$ | $L2$, $L4$ | |
|---|---|---|---|---|
| | | $l$ | $c$ | $r$ |
| $L1$, $L3$ | $u$ | - | NE | - |
| $L2$, $L4$ | $m$ | NE | - | - |
| | $d$ | - | - | NE |

and therefore $L2$ and by induction all higher order $Ln$ heuristics will prescribe the same action as $L1$. In the second case, assume that the $L1$ solutions of the two players do not lead to a PSNE, however it is necessarily true that each player's action lies in the support of a PSNE. An example is given in Table XVI where the pure strategy Nash equilibrium outcomes are denoted by NE. Let the $L1$ actions for both players be $(u,l)$ leading to a non-NE outcome. The $L2$ heuristic prescription is now $(m,c)$, $L3$ again prescribes $(u,l)$ and $L4$ returns to $(m,c)$. The intuition behind this is that since neither of the $L1$ actions are in the support of the NE at $(d,r)$, then higher order $Ln$ heuristics which are best responses to the $L1$ actions cannot possibly prescribe actions $d$ and $r$. Hence, all higher order $Ln$ heuristics must oscillate between the remaining two actions of each player, so that for all odd values of $n$, $Ln$ will prescribe the same action, and likewise for all even values of $n$.

All the independent variables in the regressions are dummy variables and therefore the magnitude of the estimated coefficients can be used for direct cross-variable comparisons of the magnitude of estimated coefficients and their economic significance. Table XVII documents the results of conditional logit regressions on parts of the dataset as distinguished by the number of PSNE in the games. In all the models, the $L2$ heuristic is always the most significant factor affecting the network's behavior as it causes the largest increase in probability of playing the associated action. The *minimax* heuristic is not statistically significant at the 5% level in the full dataset of games. In games with three PSNE only the *maximax*, *naive* and $L2$ heuristics are statistically significant. Shifting attention to the dominance variables, $dominated$ actions are always played much less often as the coefficient is negative and large in magnitude (roughly of the same order as that of the $L2$ heuristic). This result comfirms previous observations that the NNs have endogenously learned to avoid playing dominated strategies. Even more interesting is the result for best responding to an opponent's dominant strategy. The large and statistically significant values of $\beta_{it.dom.}$ imply that the networks have learned this more complicated behavior as well. The fit as measured by McFadden's $Pseudo-R^2$ is much lower in the models explaining games with either zero or three PSNE, compared to those with one or two PSNE.

As an alternative means of assessing how well these heuristics explain the behavior of the NNs, Table XVIII provides statistics on the percentage of the NN's actions correctly predicted by a single heuristic. The results confirm the main conclusion from the above conditional logit regressions as the L2 heuristic is superior at predicting the NNs' responses on the whole set of games. The next best heuristics are $L3$ and

TABLE XVII

LOGIT REGRESSION OF NN OUTPUT GROUPED BY THE NO. OF PSNE OF GAMES

| No. of PSNE | | Coef. | Std. Err. | $t$-stat. | $p$-value | $lower_{95\%}$ | $upper_{95\%}$ |
|---|---|---|---|---|---|---|---|
| $PSNE = 0$ | $\beta_{minimax}$ | 0.000 | 0.017 | -0.010 | 0.996 | -0.034 | 0.034 |
| | $\beta_{br-minimax}$ | 0.037 | 0.015 | 2.480 | 0.013 | 0.008 | 0.067 |
| | $\beta_{maximax}$ | 0.059 | 0.016 | 3.740 | 0.000 | 0.028 | 0.089 |
| | $\beta_{br-maximax}$ | 0.039 | 0.015 | 2.620 | 0.009 | 0.010 | 0.069 |
| | $\beta_{L1}$ | 0.304 | 0.029 | 10.600 | 0.000 | 0.248 | 0.360 |
| | $\beta_{L2}$ | 0.618 | 0.028 | 21.980 | 0.000 | 0.563 | 0.673 |
| | $\beta_{L3}$ | 0.274 | 0.028 | 9.860 | 0.000 | 0.220 | 0.328 |
| | $\beta_{L4}$ | 0.327 | 0.028 | 11.570 | 0.000 | 0.271 | 0.382 |
| | $\beta_{dominated}$ | -0.433 | 0.042 | -10.400 | 0.000 | -0.515 | -0.351 |
| | $LR\,\chi^2(7)$ | 3828.56 | | | | | |
| | $Pseudo-R^2$ | 0.0808 | | | | | |
| $PSNE = 1$ | $\beta_{minimax}$ | -0.015 | 0.013 | -1.180 | 0.239 | -0.040 | 0.010 |
| | $\beta_{br-minimax}$ | 0.066 | 0.011 | 5.830 | 0.000 | 0.044 | 0.088 |
| | $\beta_{maximax}$ | 0.049 | 0.012 | 4.130 | 0.000 | 0.026 | 0.072 |
| | $\beta_{br-maximax}$ | 0.086 | 0.011 | 7.660 | 0.000 | 0.064 | 0.109 |
| | $\beta_{L1}$ | 0.199 | 0.014 | 14.600 | 0.000 | 0.173 | 0.226 |
| | $\beta_{L2}$ | 0.384 | 0.015 | 25.320 | 0.000 | 0.354 | 0.414 |
| | $\beta_{L3}$ | 0.302 | 0.020 | 15.160 | 0.000 | 0.263 | 0.342 |
| | $\beta_{L4}$ | 0.507 | 0.021 | 24.260 | 0.000 | 0.466 | 0.548 |
| | $\beta_{it.dom.}$ | 0.130 | 0.024 | 5.410 | 0.000 | 0.083 | 0.177 |
| | $\beta_{dominated}$ | -0.398 | 0.022 | -17.950 | 0.000 | -0.442 | -0.355 |
| | $LR\,\chi^2(8)$ | 31178.90 | | | | | |
| | $Pseudo-R^2$ | 0.2441 | | | | | |
| $PSNE = 2$ | $\beta_{minimax}$ | 0.006 | 0.019 | 0.330 | 0.740 | -0.030 | 0.043 |
| | $\beta_{br-minimax}$ | 0.084 | 0.016 | 5.220 | 0.000 | 0.053 | 0.116 |
| | $\beta_{maximax}$ | 0.113 | 0.017 | 6.710 | 0.000 | 0.080 | 0.146 |
| | $\beta_{br-maximax}$ | 0.101 | 0.016 | 6.280 | 0.000 | 0.070 | 0.133 |
| | $\beta_{L1}$ | -0.054 | 0.031 | -1.730 | 0.084 | -0.114 | 0.007 |
| | $\beta_{L2}$ | 0.165 | 0.095 | 1.740 | 0.081 | -0.021 | 0.351 |
| | $\beta_{L3}$ | 0.451 | 0.029 | 15.380 | 0.000 | 0.394 | 0.509 |
| | $\beta_{L4}$ | 0.319 | 0.095 | 3.350 | 0.001 | 0.132 | 0.506 |
| | $\beta_{dominated}$ | -0.879 | 0.039 | -22.280 | 0.000 | -0.956 | -0.802 |
| | $LR\,\chi^2(7)$ | 5531.11 | | | | | |
| | $Pseudo-R^2$ | 0.1295 | | | | | |
| $PSNE = 3$ | $\beta_{minimax}$ | 0.082 | 0.085 | 0.970 | 0.331 | -0.084 | 0.248 |
| | $\beta_{br-minimax}$ | 0.104 | 0.073 | 1.420 | 0.155 | -0.039 | 0.247 |
| | $\beta_{maximax}$ | 0.071 | 0.075 | 0.950 | 0.341 | -0.075 | 0.217 |
| | $\beta_{br-maximax}$ | 0.079 | 0.073 | 1.080 | 0.282 | -0.065 | 0.222 |
| | $\beta_{L1}$ | 0.218 | 0.085 | 2.580 | 0.010 | 0.052 | 0.383 |
| | $\beta_{L2}$ | 0.377 | 0.070 | 5.370 | 0.000 | 0.239 | 0.515 |
| | $LR\,\chi^2(6)$ | 46.16 | | | | | |
| | $Pseudo-R^2$ | 0.0248 | | | | | |

TABLE XVIII

PERCENTAGE OF NN ACTIONS CORRECTLY PREDICTED BY HEURISTICS IN $\Psi^{3}_{hom}$

| | Heuristics | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| # PSNE | *minimax* | *br − minimax* | *maximax* | *br − maximax* | *L*1 | *L*2 | *L*3 | *L*4 |
| Any | 49.8 | 53 | 53.5 | 52.7 | 58.2 | 72.8 | 70.4 | 68.2 |
| 0 | 35 | 39.8 | 39.2 | 39.9 | 40.2 | 56 | 37.6 | 22.9 |
| 1 | 56.8 | 60.3 | 60.5 | 59.9 | 66.1 | 83.2 | 85.9 | 87.6 |
| 2 | 45.8 | 46.8 | 49.4 | 46.2 | 55.4 | 61.6 | 61.7 | 61.9 |
| 3 | 33.7 | 28.2 | 31.7 | 33.3 | 37.6 | 42.5 | 37.6 | 42.4 |

*L*4 followed by *L*1 with the rest of the heuristics performing significantly worse. It is interesting that we have found evidence of heuristics not only incorporating own payoffs, such as *minimax*, *maximax* and *L*1 but also heuristics involving best responses to opponents' heuristics and therefore opponents' payoffs. The most difficult games for prediction are those with three PSNE and then those with zero PSNE, agreeing with the conclusion from the conditional logit regressions. The most notable observation from the data on specific games is that the *L*3 and *L*4 heuristics outperform *L*2 in games with a single PSNE.

### 5.7.2. *Evolution of heuristic learning during training*

The previous analysis focused on the final trained behavior of the NNs after one million generations, but there is also the issue of how the networks behaved during the training and learning process. Indeed, there is no reason to believe that human behavior is best approximated by the final networks, but rather may be better explained by some level of experience corresponding to less than a million generations. Table XIX catalogs the results from a series of conditional logit regressions estimated at different points during the training process (data on all games regardless of number and type of NE was employed). The first column denotes the subset of generations over which each regression was estimated, with the subsets created cumulatively so that each value in this column gives the upper bound of the subset, the lower bound being defined by the upper bound of the immediately preceding subset. For example, for the row with value 22, the sample used was from the 2,000th to the 22,000th generations.

The first observation is that the $Pseudo-R^2$ increases over time so that the chosen heuristics fit the NN behavior better as learning progresses. The heuristics *minimax*, *br − minimax* and *maximax* do not change significantly or achieve prominence during learning. Throughout the training the coefficients $\beta_{dominated}$ and $\beta_{it.dom.}$ increase in magnitude with the former however always more economically significant than the latter.

An interesting result concerns the behavior of the coefficients $\beta_{L1}$ and $\beta_{L2}$ during training. Initially the *L*1 heuristic explains behavior much better than the *L*2 heuristic, this continues to hold until roughly the 100,000th generation at which time the *L*2 heuristic overtakes the *L*1 heuristic in importance. The *L*2 heuristic continues to increase in relative importance until the end of the training period where it is undeniably the most important heuristic. This behavior is reasonable since as the population comes to play the $L1$ heuristic more often the networks in their attempt to best respond to their opponents will begin to learn to best respond to this behavior. Regarding the plausibility of NNs as models of human learning, the importance of the $L1$ and $L2$ heuristics in explaining NN behavior is encouraging, since as discussed above these two heuristics are also the best predictors of human behavior in experiments. The simulations also give a possible explanation of the heterogeneity of human behavior. Use of the $L1$ heuristic may be more predominant in individuals with less experience, with human behavior evolving to the $L2$ heuristic with more experience[21].

Since the NNs try to best respond to the behavior of their opponents, a reasonable hypothesis is that the incidence of $Ln$ play, for any integer $n$, will be subject to the following life-cycle. As incidence of $Ln-1$ play increases this will eventually lead to an increase in the use of $Ln$, as a best response to $Ln-1$. As $Ln$ play becomes adopted the use of the $Ln-1$ heuristic should start declining as it will become relatively unprofitable. This evolutionary life-cycle hypothesis of the heuristics is upheld by the data in Table XIX. The $L1$ heuristic is initially increasing until the point in time when the best response to it, $L2$, reaches a critical mass and thereafter is decreasing. The $L3$ and $L4$ heuristics are initially not important in explaining NN behavior, but as the heuristic one hierarchical step lower than them becomes prevalent they begin to become more important. For the duration of this training session the $Ln$ heuristics where $n \geq 2$, do not reach the point where their use starts declining, however a reasonable conjecture is that given enough training generations they would also fall prey to this life-cycle hypothesis. It should be noted, that the $L1$ heuristic entered its declining phase very quickly, namely after 42,000 generations whilst the rest of the $Ln$ heuristics did not enter the declining phase during these 1,000,000 generations. Hence, these higher-order heuristics appear to have a longer life-cycle than less complicated heuristics. This is a reasonable conjecture because learning higher order statistics should be increasingly hard for the neural networks and therefore the rate of learning will probably slow down, thereby affording more time to the heuristics before they become obsolete from the emergence of more sophisticated heuristics. The apparent positive correlation between heuristic sophistication and the length of its life-cycle provides a practical limit to the rationality of real agents, who have a finite history of experience.

The above results pooled the networks' behavior into one equation independent of the type of game that was played. The same model is now applied to subsets of the games based on the number of PSNE of games. This allows for the possibility that the NNs have learned to classify games according to a similarity measure that arose endogenously, and therefore behave similarly for games in the same class (i.e. with the same strategic properties) but differently for games in different classes. In Table XX which provides the data for games

[21] A more rigorous test of this hypothesis would involve running simulations with birth and death of NN agents so that inexperienced individuals are suddenly thrown into a population of more experienced agents.

TABLE XIX.— Logit regressions of NN output throughout the training phase of $\Psi^3_{hom}$ for all games

| Generations | Estimated coefficients | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $(\times 10^3)$ | $\beta_{minimax}$ | $\beta_{br-minimax}$ | $\beta_{maximax}$ | $\beta_{br-maximax}$ | $\beta_{L1}$ | $\beta_{L2}$ | $\beta_{L3}$ | $\beta_{L4}$ | $\beta_{it.dom.}$ | $\beta_{dominated}$ | $Pseudo-R^2$ |
| 2 | 0.094 | 0.019 | 0.078 | 0.038 | 0.230 | -0.002 | 0.031 | 0.010 | -0.007 | -0.226 | 0.025 |
| 22 | 0.130 | 0.008 | 0.095 | 0.012 | 0.394 | 0.122 | 0.025 | 0.049 | -0.027 | -0.382 | 0.069 |
| 42 | 0.123 | 0.026 | 0.112 | 0.034 | 0.401 | 0.172 | 0.038 | 0.071 | 0.096 | -0.340 | 0.082 |
| 62 | 0.106 | 0.014 | 0.149 | 0.026 | 0.356 | 0.254 | 0.053 | 0.081 | 0.142 | -0.337 | 0.089 |
| 82 | 0.058 | 0.030 | 0.137 | 0.013 | 0.372 | 0.288 | 0.074 | 0.058 | 0.121 | -0.393 | 0.093 |
| 102 | 0.050 | 0.012 | 0.130 | 0.006 | 0.382 | 0.313 | 0.107 | 0.074 | 0.139 | -0.372 | 0.099 |
| 122 | 0.008 | 0.054 | 0.124 | 0.001 | 0.373 | 0.346 | 0.098 | 0.065 | 0.158 | -0.382 | 0.101 |
| 142 | 0.034 | 0.023 | 0.128 | 0.020 | 0.359 | 0.369 | 0.145 | 0.075 | 0.082 | -0.344 | 0.107 |
| 162 | 0.031 | 0.026 | 0.130 | 0.038 | 0.331 | 0.409 | 0.092 | 0.093 | 0.114 | -0.335 | 0.106 |
| 182 | 0.039 | 0.039 | 0.089 | 0.041 | 0.311 | 0.402 | 0.140 | 0.083 | 0.163 | -0.418 | 0.113 |
| 202 | 0.024 | 0.066 | 0.070 | 0.031 | 0.365 | 0.400 | 0.127 | 0.082 | 0.132 | -0.391 | 0.113 |
| 300 | 0.008 | 0.069 | 0.075 | 0.035 | 0.332 | 0.458 | 0.168 | 0.125 | 0.215 | -0.404 | 0.134 |
| 400 | -0.009 | 0.055 | 0.104 | 0.059 | 0.276 | 0.455 | 0.197 | 0.097 | 0.148 | -0.510 | 0.136 |
| 500 | -0.012 | 0.068 | 0.068 | 0.057 | 0.237 | 0.473 | 0.262 | 0.118 | 0.115 | -0.572 | 0.149 |
| 600 | 0.022 | 0.054 | 0.069 | 0.094 | 0.240 | 0.463 | 0.253 | 0.177 | 0.178 | -0.461 | 0.153 |
| 700 | -0.012 | 0.073 | 0.071 | 0.071 | 0.244 | 0.434 | 0.310 | 0.206 | 0.091 | -0.540 | 0.164 |
| 800 | 0.010 | 0.028 | 0.067 | 0.077 | 0.220 | 0.442 | 0.275 | 0.215 | 0.211 | -0.558 | 0.161 |
| 900 | -0.012 | 0.065 | 0.057 | 0.059 | 0.217 | 0.437 | 0.325 | 0.246 | 0.222 | -0.600 | 0.180 |
| 1000 | -0.030 | 0.056 | 0.047 | 0.077 | 0.254 | 0.450 | 0.360 | 0.230 | 0.178 | -0.509 | 0.176 |

with no PSNE, the pattern is similar to the general results discussed above. Initially, the $L1$ heuristic better explains the data but as the networks gain experience, the $L2$ heuristic becomes more important in explaining behavior. Also, actions that are dominated are played significantly less often. In contrast to the results from the analysis for all games, the $L2$ heuristic clearly retains its status as the most important heuristic till the end of the training, as its estimated coefficient is almost double that of the next most important heuristic. Also, in this case the $L1$ heuristic is more important than the $L3$ heuristic throughout the training.

Shifting focus to games with a unique PSNE, whose results are documented in Table XXI, the rank ordering of importance of the heuristics changes appreciably. Initially, at the early stages of training $L1$ is again the most important variable, only to be surpassed in due time by the $L2$ heuristic. All the while $\beta_{L3}$ is increasing over time, overtaking $\beta_{L1}$ in magnitude between 600,000 and 700,000 generations, but never surpassing $\beta_{L2}$ in magnitude. The most striking difference compared to previous results is that the $L4$ heuristic achieves much greater prominence, overtaking even the $L2$ heuristic after roughly 800,000-900,000 generations. The fit of the models as measured by the $Pseudo-R^2$ is also much higher than for the other subsets of games investigated.

Table XXII reveals another importance difference in NN behavior when the games played have two PSNE. Although as usual the $L1$ heuristic is initially an important variable, by the end of the training $\beta_{L1}$ is approximately equal to zero. The largest in magnitude coefficient is $\beta_{L3}$ by the end of the training period, with $\beta_{L2}$ and $\beta_{L4}$ following. The most striking result for this subgroup of games is the variability in the estimated coefficient throughout the NN training. Table XXIII also attests to significant variability in the regressions for games with three PSNE.

The divergence of the parameter estimates of these models for games with different numbers of PSNE is strong evidence that the NNs have learnt to classify games according to their strategic properties. This is impressive as the networks were not preprogrammed with any measure of similarity of games, but nonetheless a fairly sophisticated similarity measure has arisen endogenously from experience.

#### 5.7.3. *Heuristics of bounded rational NNs*

The repercussions of varying the sophistication of the topologies of the NNs on their reasoning and behavior is examined in Table XXIV. An evident trend exists for more sophisticated networks to use higher order $Ln$ heuristics, such as $L3$ and $L4$, and also to conform more to the prescriptions of dominance and iterated dominance. $L1$ is the most important heuristic as gauged by the magnitude of the estimated coefficients, for the networks $\psi_1\{\varnothing,\varnothing\}$, $\psi_2\{5,1\}$, $\psi_3\{5,2\}$ and $\psi_4\{5,3\}$ i.e. the four simplest networks. The $L2$ heuristic then comes to dominate the behavior of the remaining networks, with the $L1$ and $L3$ heuristics trailing next in importance for the most sophisticated networks. The finding that for the most sophisticated networks

TABLE XX.— Logit regressions of NN output throughout the training phase of $\Psi^3_{hom}$ for games with zero PSNE

| Generations | Estimated coefficients | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ($\times 10^3$) | $\beta_{minimax}$ | $\beta_{br-minimax}$ | $\beta_{maximax}$ | $\beta_{br-maximax}$ | $\beta_{L1}$ | $\beta_{L2}$ | $\beta_{L3}$ | $\beta_{L4}$ | $\beta_{it.dom.}$ | $\beta_{dominated}$ | $Pseudo-R^2$ |
| 2 | 0.081 | 0.058 | 0.085 | 0.038 | 0.290 | 0.046 | 0.077 | 0.062 | - | -0.067 | 0.020 |
| 22 | 0.130 | 0.037 | 0.118 | -0.032 | 0.522 | 0.120 | 0.114 | 0.029 | - | -0.433 | 0.068 |
| 42 | 0.133 | 0.074 | 0.097 | -0.019 | 0.468 | 0.210 | 0.033 | 0.132 | - | -0.184 | 0.058 |
| 62 | 0.062 | -0.077 | 0.095 | 0.022 | 0.356 | 0.378 | -0.055 | 0.194 | - | -0.347 | 0.056 |
| 82 | 0.037 | 0.038 | 0.067 | -0.035 | 0.480 | 0.358 | 0.080 | 0.106 | - | -0.328 | 0.063 |
| 102 | 0.027 | 0.015 | 0.095 | -0.014 | 0.483 | 0.385 | 0.130 | 0.144 | - | -0.347 | 0.066 |
| 122 | -0.004 | 0.054 | 0.167 | -0.021 | 0.390 | 0.412 | 0.046 | 0.110 | - | -0.423 | 0.071 |
| 142 | 0.025 | 0.039 | 0.154 | -0.030 | 0.506 | 0.439 | 0.217 | 0.105 | - | -0.297 | 0.077 |
| 162 | 0.006 | -0.020 | 0.154 | -0.077 | 0.416 | 0.399 | 0.110 | 0.038 | - | -0.461 | 0.067 |
| 182 | 0.007 | -0.067 | 0.080 | 0.018 | 0.484 | 0.470 | 0.206 | 0.087 | - | -0.402 | 0.074 |
| 202 | 0.054 | -0.012 | 0.054 | 0.006 | 0.319 | 0.595 | 0.001 | 0.247 | - | -0.379 | 0.072 |
| 300 | -0.011 | -0.043 | 0.036 | 0.002 | 0.452 | 0.620 | 0.157 | 0.219 | - | -0.369 | 0.081 |
| 400 | 0.007 | 0.036 | 0.084 | 0.015 | 0.466 | 0.548 | 0.296 | 0.153 | - | -0.339 | 0.083 |
| 500 | -0.017 | 0.065 | 0.016 | 0.012 | 0.360 | 0.680 | 0.261 | 0.264 | - | -0.477 | 0.093 |
| 600 | 0.077 | -0.052 | 0.087 | 0.046 | 0.211 | 0.764 | 0.166 | 0.424 | - | -0.382 | 0.085 |
| 700 | -0.021 | 0.049 | 0.052 | 0.051 | 0.319 | 0.524 | 0.276 | 0.245 | - | -0.520 | 0.076 |
| 800 | -0.009 | -0.052 | 0.011 | 0.001 | 0.303 | 0.766 | 0.157 | 0.460 | - | -0.242 | 0.073 |
| 900 | -0.037 | 0.000 | 0.004 | -0.022 | 0.282 | 0.680 | 0.244 | 0.404 | - | -0.595 | 0.085 |
| 1000 | -0.037 | -0.003 | 0.011 | 0.052 | 0.367 | 0.643 | 0.329 | 0.322 | - | -0.446 | 0.085 |

TABLE XXI.— Logit regressions of NN output throughout the training phase of $\Psi^3_{hom}$ for games with one PSNE

| Generations | Estimated coefficients | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $(\times 10^3)$ | $\beta_{minimax}$ | $\beta_{br-minimax}$ | $\beta_{maximax}$ | $\beta_{br-maximax}$ | $\beta_{L1}$ | $\beta_{L2}$ | $\beta_{L3}$ | $\beta_{L4}$ | $\beta_{it.dom.}$ | $\beta_{dominated}$ | $Pseudo-R^2$ |
| 2 | 0.095 | 0.015 | 0.097 | 0.010 | 0.227 | 0.003 | 0.004 | 0.048 | -0.024 | -0.251 | 0.031 |
| 22 | 0.131 | 0.001 | 0.064 | 0.013 | 0.386 | 0.087 | 0.077 | 0.064 | -0.039 | -0.355 | 0.077 |
| 42 | 0.103 | 0.030 | 0.097 | 0.044 | 0.377 | 0.226 | 0.059 | 0.040 | 0.060 | -0.352 | 0.097 |
| 62 | 0.112 | 0.045 | 0.162 | 0.039 | 0.339 | 0.254 | 0.094 | 0.072 | 0.118 | -0.252 | 0.108 |
| 82 | 0.039 | 0.015 | 0.148 | 0.029 | 0.348 | 0.264 | 0.079 | 0.127 | 0.083 | -0.372 | 0.113 |
| 102 | 0.060 | 0.010 | 0.122 | -0.002 | 0.362 | 0.317 | 0.110 | 0.098 | 0.119 | -0.363 | 0.121 |
| 122 | 0.004 | 0.043 | 0.123 | -0.005 | 0.362 | 0.330 | 0.126 | 0.115 | 0.122 | -0.320 | 0.122 |
| 142 | 0.038 | 0.014 | 0.081 | 0.025 | 0.350 | 0.301 | 0.171 | 0.168 | 0.047 | -0.339 | 0.130 |
| 162 | 0.032 | 0.030 | 0.093 | 0.089 | 0.347 | 0.384 | 0.055 | 0.206 | 0.076 | -0.252 | 0.133 |
| 182 | 0.045 | 0.058 | 0.080 | 0.039 | 0.295 | 0.339 | 0.110 | 0.241 | 0.117 | -0.346 | 0.142 |
| 202 | 0.023 | 0.091 | 0.082 | 0.033 | 0.337 | 0.360 | 0.131 | 0.142 | 0.117 | -0.346 | 0.138 |
| 300 | 0.021 | 0.102 | 0.071 | 0.025 | 0.291 | 0.383 | 0.159 | 0.278 | 0.170 | -0.318 | 0.167 |
| 400 | -0.018 | 0.065 | 0.067 | 0.062 | 0.276 | 0.426 | 0.187 | 0.213 | 0.091 | -0.451 | 0.171 |
| 500 | 0.014 | 0.057 | 0.090 | 0.068 | 0.189 | 0.392 | 0.261 | 0.279 | 0.056 | -0.481 | 0.189 |
| 600 | -0.001 | 0.064 | 0.047 | 0.087 | 0.251 | 0.426 | 0.221 | 0.339 | 0.115 | -0.366 | 0.197 |
| 700 | -0.002 | 0.106 | 0.036 | 0.097 | 0.236 | 0.406 | 0.338 | 0.301 | 0.023 | -0.395 | 0.212 |
| 800 | 0.007 | 0.033 | 0.036 | 0.087 | 0.186 | 0.371 | 0.342 | 0.356 | 0.098 | -0.505 | 0.218 |
| 900 | 0.008 | 0.063 | 0.057 | 0.072 | 0.187 | 0.357 | 0.294 | 0.476 | 0.136 | -0.454 | 0.237 |
| 1000 | -0.021 | 0.077 | 0.048 | 0.061 | 0.206 | 0.345 | 0.281 | 0.546 | 0.088 | -0.371 | 0.233 |

TABLE XXII.— Logit regressions of NN output throughout the training phase of $\Psi^3_{hom}$ for games with two PSNE

| Generations | Estimated coefficients | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $(\times 10^3)$ | $\beta_{minimax}$ | $\beta_{br-minimax}$ | $\beta_{maximax}$ | $\beta_{br-maximax}$ | $\beta_{L1}$ | $\beta_{L2}$ | $\beta_{L3}$ | $\beta_{L4}$ | $\beta_{it.dom.}$ | $\beta_{dominated}$ | $Pseudo-R^2$ |
| 2 | 0.092 | -0.024 | 0.010 | 0.081 | 0.146 | 0.262 | 0.120 | -0.317 | - | -0.237 | 0.018 |
| 22 | 0.149 | -0.010 | 0.152 | 0.046 | 0.282 | 0.326 | 0.064 | -0.186 | - | -0.392 | 0.054 |
| 42 | 0.155 | -0.050 | 0.158 | 0.068 | 0.285 | -0.043 | 0.140 | 0.193 | - | -0.422 | 0.067 |
| 62 | 0.142 | 0.023 | 0.171 | -0.003 | 0.238 | 0.292 | 0.117 | -0.028 | - | -0.545 | 0.073 |
| 82 | 0.113 | 0.047 | 0.178 | 0.013 | 0.238 | 0.508 | 0.137 | -0.240 | - | -0.476 | 0.074 |
| 102 | 0.064 | 0.010 | 0.177 | 0.022 | 0.226 | 0.206 | 0.231 | 0.096 | - | -0.361 | 0.077 |
| 122 | 0.020 | 0.073 | 0.088 | 0.011 | 0.264 | 0.448 | 0.178 | -0.118 | - | -0.544 | 0.080 |
| 142 | 0.028 | 0.005 | 0.167 | 0.046 | 0.178 | 0.715 | 0.264 | -0.361 | - | -0.348 | 0.079 |
| 162 | 0.053 | 0.045 | 0.176 | 0.036 | 0.112 | 0.229 | 0.203 | 0.163 | - | -0.441 | 0.074 |
| 182 | 0.049 | 0.075 | 0.083 | 0.062 | 0.125 | 0.174 | 0.238 | 0.194 | - | -0.568 | 0.082 |
| 202 | -0.011 | 0.062 | 0.050 | 0.048 | 0.250 | 0.601 | 0.268 | -0.182 | - | -0.490 | 0.096 |
| 300 | -0.001 | 0.108 | 0.113 | 0.069 | 0.200 | 0.245 | 0.226 | 0.209 | - | -0.611 | 0.106 |
| 400 | -0.001 | 0.038 | 0.186 | 0.079 | 0.008 | 0.104 | 0.345 | 0.304 | - | -0.734 | 0.103 |
| 500 | -0.089 | 0.078 | 0.038 | 0.048 | 0.129 | 0.497 | 0.349 | -0.072 | - | -0.795 | 0.107 |
| 600 | 0.027 | 0.136 | 0.051 | 0.130 | 0.126 | 0.399 | 0.270 | 0.050 | - | -0.710 | 0.112 |
| 700 | -0.012 | 0.020 | 0.133 | 0.023 | 0.014 | 0.025 | 0.426 | 0.456 | - | -0.957 | 0.127 |
| 800 | 0.025 | 0.047 | 0.132 | 0.107 | -0.018 | 0.039 | 0.336 | 0.385 | - | -0.871 | 0.108 |
| 900 | -0.027 | 0.111 | 0.089 | 0.085 | 0.049 | 0.365 | 0.376 | 0.109 | - | -0.930 | 0.130 |
| 1000 | -0.054 | 0.052 | 0.058 | 0.107 | 0.040 | 0.260 | 0.490 | 0.197 | - | -0.942 | 0.135 |

TABLE XXIII.— Logit regressions of NN output throughout the training phase of $\Psi^3_{hom}$ for games with three PSNE

| Generations | Estimated coefficients | | | | | | |
|---|---|---|---|---|---|---|---|
| $(\times 10^3)$ | $\beta_{minimax}$ | $\beta_{br-minimax}$ | $\beta_{maximax}$ | $\beta_{br-maximax}$ | $\beta_{L1}$ | $\beta_{L2}$ | $Pseudo-R^2$ |
| 2 | 0.386 | -0.064 | 0.063 | 0.063 | 0.035 | 0.018 | 0.019 |
| 22 | -0.024 | -0.069 | -0.114 | 0.142 | 0.629 | 0.105 | 0.046 |
| 42 | 0.113 | -0.062 | -0.088 | -0.269 | 0.433 | 0.386 | 0.055 |
| 62 | 0.057 | 0.039 | 0.229 | -0.130 | 0.558 | 0.359 | 0.046 |
| 82 | 0.160 | -0.183 | 0.073 | -0.142 | 0.455 | 0.033 | 0.040 |
| 102 | -0.139 | -0.211 | 0.071 | 0.074 | 0.495 | 0.384 | 0.040 |
| 122 | 0.248 | -0.006 | -0.133 | 0.258 | 0.227 | 0.149 | 0.030 |
| 142 | 0.158 | -0.016 | 0.474 | -0.009 | 0.248 | 0.416 | 0.053 |
| 162 | -0.151 | 0.153 | 0.228 | 0.087 | 0.570 | 0.549 | 0.077 |
| 182 | 0.193 | 0.275 | 0.327 | -0.010 | 0.180 | 0.242 | 0.040 |
| 202 | 0.170 | 0.346 | -0.085 | -0.121 | 0.266 | 0.006 | 0.034 |
| 300 | 0.089 | -0.076 | -0.050 | 0.234 | 0.344 | 0.487 | 0.050 |
| 400 | 0.101 | -0.311 | 0.269 | -0.166 | 0.136 | -0.046 | 0.021 |
| 500 | 0.448 | -0.158 | 0.299 | 0.079 | -0.051 | 0.572 | 0.073 |
| 600 | -0.249 | -0.064 | 0.211 | 0.108 | 0.410 | 0.559 | 0.052 |
| 700 | -0.207 | -0.098 | 0.430 | -0.083 | 0.788 | 0.644 | 0.112 |
| 800 | 0.328 | 0.144 | 0.658 | -0.183 | 0.187 | 0.410 | 0.074 |
| 850 | 0.136 | 0.205 | 0.278 | 0.207 | 0.294 | 0.489 | 0.058 |
| 1000 | 0.190 | -0.102 | 0.091 | 0.133 | 0.229 | 0.158 | 0.021 |

TABLE XXIV

LOGIT REGRESSION OF NN OUTPUT IN THE $\Psi^3_{het}$ SIMULATION

| | NN agents | | | | |
|---|---|---|---|---|---|
| | $\psi_1\{\varnothing,\varnothing\}$ | $\psi_2\{5,1\}$ | $\psi_3\{5,2\}$ | $\psi_4\{5,3\}$ | $\psi_5\{20,1\}$ |
| $\beta_{minimax}$ | 0.190 | 0.093 | 0.117 | 0.077 | 0.003 |
| $\beta_{br-minimax}$ | 0.039 | 0.028 | 0.030 | -0.004 | 0.058 |
| $\beta_{maximax}$ | 0.121 | 0.119 | 0.116 | 0.134 | 0.069 |
| $\beta_{br-maximax}$ | 0.059 | 0.033 | -0.012 | 0.058 | 0.073 |
| $\beta_{L1}$ | 0.432 | 0.341 | 0.336 | 0.352 | 0.361 |
| $\beta_{L2}$ | 0.031 | 0.228 | 0.219 | 0.221 | 0.407 |
| $\beta_{L3}$ | -0.007 | 0.038 | 0.136 | 0.019 | 0.131 |
| $\beta_{L4}$ | 0.041 | 0.023 | 0.021 | 0.047 | 0.075 |
| $\beta_{it.dom.}$ | -0.073 | 0.074 | -0.001 | 0.098 | 0.143 |
| $\beta_{dominated}$ | -0.388 | -0.479 | -0.396 | -0.493 | -0.391 |
| $Pseudo-R^2$ | 0.079 | 0.082 | 0.081 | 0.084 | 0.114 |
| | NN agents | | | | |
| | $\psi_6\{20,2\}$ | $\psi_7\{20,3\}$ | $\psi_8\{50,1\}$ | $\psi_9\{50,2\}$ | $\psi_{10}\{50,3\}$ |
| $\beta_{minimax}$ | 0.006 | 0.013 | 0.007 | -0.040 | -0.014 |
| $\beta_{br-minimax}$ | 0.172 | 0.083 | 0.045 | 0.078 | 0.080 |
| $\beta_{maximax}$ | 0.078 | 0.094 | 0.078 | 0.071 | 0.064 |
| $\beta_{br-maximax}$ | 0.075 | 0.075 | 0.042 | 0.118 | 0.070 |
| $\beta_{L1}$ | 0.234 | 0.229 | 0.337 | 0.271 | 0.219 |
| $\beta_{L2}$ | 0.462 | 0.524 | 0.483 | 0.533 | 0.502 |
| $\beta_{L3}$ | 0.150 | 0.141 | 0.052 | 0.204 | 0.235 |
| $\beta_{L4}$ | 0.113 | 0.104 | 0.087 | 0.094 | 0.099 |
| $\beta_{it.dom.}$ | 0.148 | 0.209 | 0.176 | 0.138 | 0.224 |
| $\beta_{dominated}$ | -0.450 | -0.573 | -0.451 | -0.509 | -0.564 |
| $Pseudo-R^2$ | 0.134 | 0.147 | 0.118 | 0.154 | 0.149 |

the $\beta_{minimax}$, $\beta_{br-minimax}$, $\beta_{maximax}$ and $\beta_{br-maximax}$ coefficients are of much smaller magnitude compared to the estimated coefficients of the other heuristics also deserves mention. However, in the simple linear NN $\psi_1\{\varnothing,\varnothing\}$, both $\beta_{minimax}$ and $\beta_{maximax}$ are the most economically significant variables after $L1$. The fit of the models tends to increase with the sophistication of the agents as exemplified by the $pseudo-R^2$ statistics provided in the table.

Another avenue for investigating the heuristics which fit the networks' behavior best is a simple tabulation of the percentage of the NNs' chosen actions that are correctly predicted by each heuristic. Table XXV gives the results for all games and for sets of games broken down by number of PSNE. For all possible types of games the most successful predictions are given by the $L2$ heuristic, however it should be noted that the success rate is not much higher than that for all the other $Ln$ heuristics. In games with no PSNE, or mixed strategy games, $L1$ and $L2$ make the correct predictions 51-52% of the time, significantly outperforming all other heuristics. Shifting attention to games with a single PSNE reveals that the $L2$, $L3$ and $L4$ heuristics all predict with an almost equal success rate of 70-71%, outperforming all other heuristics. The $L3$ heuristic

TABLE XXV

PERCENTAGE OF NN ACTIONS CORRECTLY PREDICTED BY HEURISTICS IN $\Psi^3_{het}$

| | Heuristics | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| # PSNE | *minimax* | *br − minimax* | *maximax* | *br − maximax* | *L*1 | *L*2 | *L*3 | *L*4 |
| Any | 50.2 | 49.4 | 52 | 49 | 60.7 | 63.0 | 57.4 | 56.9 |
| 0 | 41.3 | 39.5 | 43 | 39.5 | 51.4 | 51.6 | 23.3 | 22.7 |
| 1 | 55.8 | 55.7 | 57.7 | 55.4 | 66.2 | 70.9 | 70.3 | 71 |
| 2 | 44 | 42.7 | 45.8 | 41.5 | 55.4 | 53.4 | 57.6 | 53.4 |
| 3 | 30.4 | 21.3 | 25.7 | 25.6 | 37.3 | 35.1 | 37.3 | 35.1 |

performs best in the games with two PSNE although the other *Ln* heuristics are not far behind. Finally, in games with three PSNE all the *Ln* heuristics have an accuracy of 35-37%, significantly above that of the other heuristics.

### 5.8. *Mixed strategy Nash equilibrium behavior of networks*

The behavior of the NNs in the various simulations when facing games with a unique MSNE is examined below[22]. One measure of whether the networks have learned to play according to the MSNE prescription to some degree is to calculate a Spearman rank correlation between the responses of the NNs and the MSNE solutions to each game. A non-parametric Spearman correlation test, $\rho$, was chosen as the probability of play is not normally distributed and it is bound between zero and one, thereby invalidating the assumptions of the standard correlation coefficient. These results are based on test datasets and not on the training dataset.

#### 5.8.1. *Analysis of NN behavior in games with a unique MSNE in* $\Psi^3_{hom}$

A Spearman correlation coefficient calculated between the NNs' output and the PSNE prescription, $\rho$, gives a high value of 0.6236. However, this calculation includes actions that are played with zero probability in a MSNE and as a consequence must include dominated actions. Hence, it is very likely that the neural networks will assign a very low probability to such an action as a consequence of the NNs ability to learn to avoid playing dominated actions. Not excluding such actions may lead to an overestimation of the degree with which the NNs have learned the MSNE. The Spearman correlation for actions in the support of the MSNE, $\rho_s$, falls to a value of 0.3926, which however is still significant.

Another approach is to model the responses of the NNs by conditional logit regressions (McFadden et al., 1973) using three different sets of independent variables. In model 1, only heuristics are included as regressors, in model 2 only the MSNE solution is used and finally model 3 nests both of the previous models by incorporating all of the heuristics and the MSNE solution simultaneously. The results of regressing these

[22]Calculating the MSNE for $2 \times 2$ games simply involves solving two simultaneous equations but becomes much more complicated for $3 \times 3$ games. A detailed discussion of the method employed to calculate the MSNE in these games is given in Appendix C.1.

TABLE XXVI

MODELING OF NN BEHAVIOR IN GAMES WITH A UNIQUE MSNE AND NO PSNE

| | | Coefficient | Std err. | $t$-stat. | $p$-value | $lower_{95\%}$ | $upper_{95\%}$ |
|---|---|---|---|---|---|---|---|
| Model 1 | $\beta_{minimax}$ | -0.020 | 0.017 | -1.180 | 0.239 | -0.054 | 0.013 |
| | $\beta_{br-minimax}$ | 0.032 | 0.015 | 2.170 | 0.030 | 0.003 | 0.062 |
| | $\beta_{maximax}$ | 0.067 | 0.016 | 4.290 | 0.000 | 0.036 | 0.098 |
| | $\beta_{br-maximax}$ | 0.033 | 0.015 | 2.210 | 0.027 | 0.004 | 0.062 |
| | $\beta_{naive}$ | 0.287 | 0.026 | 10.890 | 0.000 | 0.236 | 0.339 |
| | $\beta_{L2}$ | 0.591 | 0.025 | 23.160 | 0.000 | 0.541 | 0.641 |
| | $\beta_{L3}$ | 0.242 | 0.025 | 9.570 | 0.000 | 0.192 | 0.291 |
| | $\beta_{L4}$ | 0.291 | 0.026 | 11.290 | 0.000 | 0.240 | 0.341 |
| | $\beta_{dominated}$ | -0.461 | 0.043 | -10.840 | 0.000 | -0.545 | -0.378 |
| | $Pseudo-R^2$ | 0.0712 | | | | | |
| Model 2 | $\beta_{msne}$ | 0.990 | 0.022 | 44.150 | 0.000 | 0.946 | 1.034 |
| | $Pseudo-R^2$ | 0.0416 | | | | | |
| Model 3 | $\beta_{minimax}$ | -0.029 | 0.017 | -1.700 | 0.088 | -0.063 | 0.004 |
| | $\beta_{br-minimax}$ | 0.034 | 0.015 | 2.300 | 0.021 | 0.005 | 0.064 |
| | $\beta_{maximax}$ | 0.070 | 0.016 | 4.460 | 0.000 | 0.039 | 0.101 |
| | $\beta_{br-maximax}$ | 0.036 | 0.015 | 2.400 | 0.016 | 0.007 | 0.065 |
| | $\beta_{naive}$ | 0.252 | 0.027 | 9.470 | 0.000 | 0.200 | 0.305 |
| | $\beta_{L2}$ | 0.491 | 0.027 | 17.890 | 0.000 | 0.437 | 0.545 |
| | $\beta_{L3}$ | 0.209 | 0.026 | 8.190 | 0.000 | 0.159 | 0.259 |
| | $\beta_{L4}$ | 0.264 | 0.026 | 10.190 | 0.000 | 0.214 | 0.315 |
| | $\beta_{dominated}$ | -0.407 | 0.043 | -9.480 | 0.000 | -0.491 | -0.323 |
| | $\beta_{msne}$ | 0.298 | 0.030 | 9.940 | 0.000 | 0.239 | 0.357 |
| | $Pseudo-R^2$ | 0.0733 | | | | | |

models are shown in Table XXVI. The measure of fit used is the $Pseudo-R^2$ proposed by McFadden et al. (1973) which is simply $1-ll(full\,model)/ll(constant\,only\,model)$, where $ll$ denotes the log-likelihood. Typically, values of this measure of fit are much lower than those of standard $R^2$, and Louviere et al. (2000) and Hensher and Johnson (1981) argue that values between 0.2 and 0.4 represent a good fit.

In model 1 where the dependent variable is explained only by standard heuristics, it is the $L2$ heuristic which stands out as the most important determinant with most of the other heuristics following at an appreciable distance. The $Pseudo-R^2$ value of 0.0712 declares that although the fit is appreciable there is still much variance to be explained. Model 2 is a much more parsimonious model as it has only one regressor compared to the heuristics used in model 1. The $R^2$ is lower than that of model 1 however it uses only the MSNE solution to explain behavior. Model 3 utilizes all the regressors from models 1 and 2 and as expected provides the best fit out of the three with a $Pseudo-R^2$ of 0.0733 although only marginally so compared to model 1. The most important heuristic is still $L2$, the coefficient on the MSNE solution is much smaller than in model 2 but still important in explaining behavior.

A density plot of actual neural network behavior, in terms of the probability of playing each action in the

FIGURE 5.— Density plot of MSNE and NN output for $3 \times 3$ games in $\Psi^{10,3,r}_{3\times 50}$

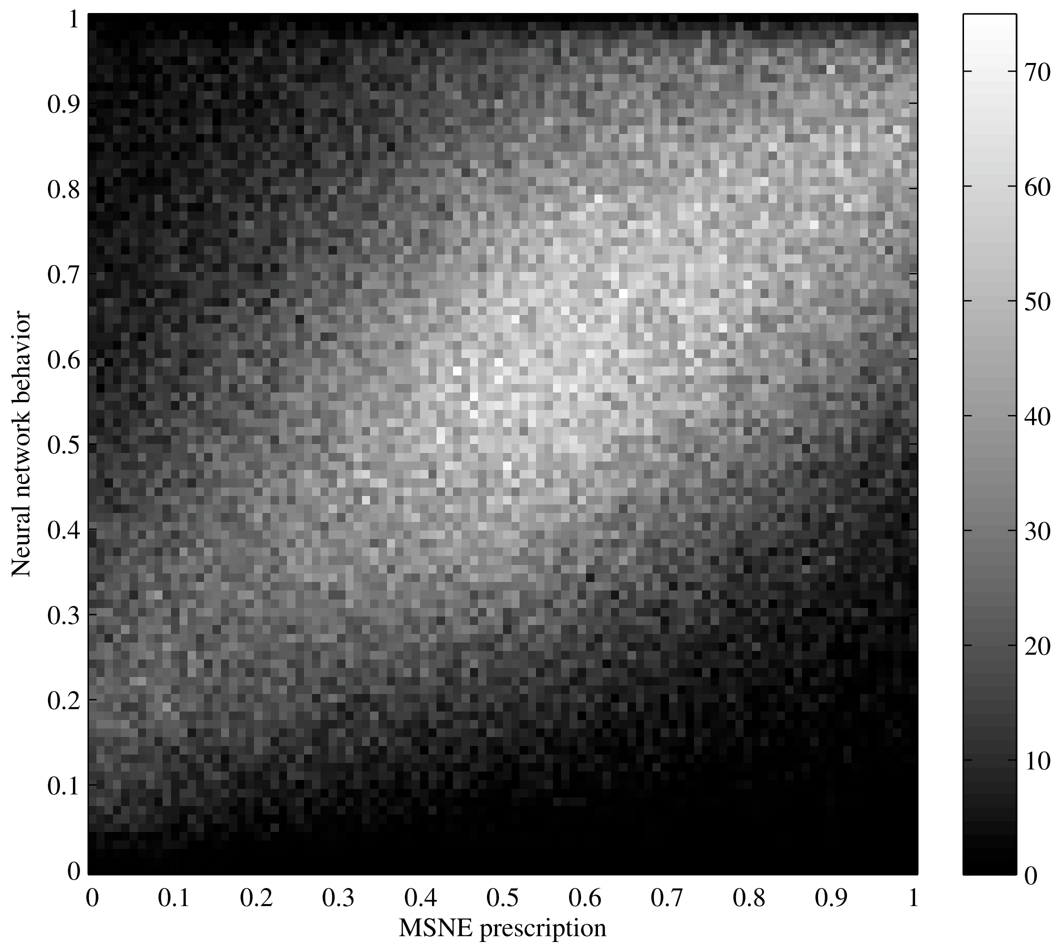


TABLE XXVII
CORRELATION OF INDIVIDUAL NNS' OUTPUT AND THE MSNE PRESCRIPTION IN $\Psi^3_{het}$

| | NN agents | | | | |
|---|---|---|---|---|---|
| Correlation | $\psi_1\{\varnothing,\varnothing\}$ | $\psi_2\{5,1\}$ | $\psi_3\{5,2\}$ | $\psi_4\{5,3\}$ | $\psi_5\{20,1\}$ |
| $\rho$ | 0.3983 | 0.4602 | 0.466 | 0.4684 | 0.555 |
| $\rho_s$ | 0.0847 | 0.1843 | 0.1865 | 0.1895 | 0.3129 |
| | NN agents | | | | |
| Correlation | $\psi_6\{20,2\}$ | $\psi_7\{20,3\}$ | $\psi_8\{50,1\}$ | $\psi_9\{50,2\}$ | $\psi_{10}\{50,3\}$ |
| $\rho$ | 0.5976 | 0.5874 | 0.5558 | 0.6103 | 0.6189 |
| $\rho_s$ | 0.3813 | 0.3576 | 0.3245 | 0.4002 | 0.3919 |

support of the MSNE, versus the MSNE prescribed probabilities from a test data set of $3 \times 3$ games with unique MSNE is presented in Figure 5. The level of shading respresents the frequency of combinations of these points, so that lighter shades represent a higher concentration of datapoints in that area. The density plot displays a roughly linear relationship throughout all values of the MSNE prescription, albeit with a large degree of dispersion around the 45° line from the origin. This confirms the statistically and economically significant $\beta_{msne}$ estimates from the conditional logit regressions.

### 5.8.2. *Analysis of NN behavior in games with a unique MSNE in* $\Psi^3_{het}$

An analysis of the Spearman correlations between types of NNs in the $\Psi^3_{het}$ simulation follows. The first row in Table XXVII compiles the Spearman correlation coefficient, $\rho$, for each type of NN including all possible actions, even those not in the support of the MSNE solution. The second row calculates the Spearman correlation only for actions in the support of the MSNE, denoted by $\rho_s$. It is clear from the table that including these observations does lead to an overestimation of the NNs' MSNE behavior as $\rho_s$ is much smaller than $\rho$ for all types of NNs.

By far, the smallest value of $\rho_s$ was exhibited by the simplest network, $\psi_1\{\varnothing,\varnothing\}$, with a value of only 0.0847, indicating that play was driven mostly by other considerations. Another observation is that $\rho_s$ is positively related to the number of neurons in each layer, but not correlated with the actual number of layers. Increasing the number of neurons per layer from 5 to 20 leads to an abrupt increase in the correlation, whereas increasing it from 20 to 50 leads to a very small increase only. The correlation for $\psi_{10}\{50,3\}$ is very close to that of the networks in the homogeneous simulation, which share the same topology. Hence, playing against other less sophisticated NNs in the heterogeneous simulation does not seem to significantly affect MSNE behavior.

# 6. COMPARISON OF NN BEHAVIOR TO HUMAN SUBJECT EXPERIMENTAL STUDIES

The previous section documented the behavior of the simulated NNs and compared it to some empirical facts from various experiments, such as PSNE play, equilibrium selection, and found a high degree of similarity. This section will go further by using five experimental studies where human subjects were presented with a variety of one-shot games. Data from these studies provide the probability distribution over actions for each game that was played by a population of subjects. Likewise, it is possible to attain the probability distribution data from the simulated NNs and then compare the results to identify how well the NNs fit the behavior of human subjects. The measure of fit that will be used is the Spearman correlation of experimentally observed probabilities of actions and the simulated probabilities of actions from the NNs.

There are two important issues that need to be addressed before this comparison can be made. Firstly, the NNs were trained on games that were necessarily bound between -1 and 1, and therefore a mechanism is required that translates payoffs from the experimental games to payoffs that are admissible to the NNs. The most intuitive way of achieving this without affecting the ranking of payoffs is simply to scale or normalize a game's payoffs so that its maximum payoff is equal to one. A more complicated approach might try to translate the game payoffs to the equivalent payoffs in real currency to the subjects, thereby keeping information about the relative magnitude of incentives for different games and experiments. However, this is extremely complicated and although it would preserve the relative payoffs there would still be an issue of how to calibrate the absolute payoffs. As a solution, all statistics will be run for different payoff scaling factors. Firstly, all games will be normalized so that the absolute value of the largest in magnitude payoff is equal to one. Then a set of these games, scaled by values between zero and one inclusive in steps of 0.1, will be created and the Spearman correlation coefficient will be estimated for each of these scaled games.

A more important issue that needs to be addressed is what level of NN experience accurately reflects the experience brought by subjects to experiments. There is no reason to believe that any arbitrary number of generations or the limiting behavior of the NNs is appropriate. Hence, the number of generations of experience is a variable that needs to be estimated. The way that this is accomplished is by estimating all statistics for different levels of experience and then observing at which level the best fit is achieved. This is especially important because the proportion of different types of games that arises with the random sampling scheme will not coincide with the proportions that have occurred in subjects' lives. Hence, a different number of generations may be necessary for each type of game to correct for this.

Essentially, two unknown parameters must be estimated or fitted, the scaling parameter for payoffs and the number of elapsed generations. The data was drawn from Ivanov (2006), Stahl and Wilson (1995), Binmore et al. (2001), Rey Biel (2004) and Tang (2001) which utilize a variety of $3 \times 3$ games with different properties.

TABLE XXVIII
NN BEHAVIOR COMPARED TO HUMAN EXPERIMENTAL STUDIES

| | Unique PSNE | | | | MSNE | | | 3 PSNE |
|---|---|---|---|---|---|---|---|---|
| | Biel | Ivanov | S&W | Binmore | Tang | S&W | Binmore | S&W |
| $\rho$ | 0.74 | 0.8165 | 0.58 | 0.8857 | 0.5121 | 0.49 | 0.7762 | 0.55 |
| Generation | 100,000 | 1,000,000 | 10,000 | 10,000 | 1,000 | 5,000 | 800,000 | 90,000 |
| Scale factor | 1 | 1 | 0.2 | 1 | 1 | 0.2 | 1 | 0.1 |

Rey Biel (2004) presented ten games to subjects, all of which exhibited a unique PSNE but differed in regards to how many, if any, levels of iterated dominance were necessary to solve the games. Ivanov (2006) also presents subjects with ten games with a unique PSNE, only three of which are dominance solvable. Binmore et al. (2001) investigates the MSNE behavior of subjects by presenting them with seven different zero-sum games, only three of which are $3 \times 3$ and will be used here. Tang (2001) uses three games which all exhibit the same MSNE, with one of the games also exhibiting two other equilibria where only one player is mixing. Stahl and Wilson (1995), in contrast to the previous studies use a set of symmetric $3 \times 3$ games with much greater variety than the previous studies. Out of the twelve games used three had a unique MSNE, five had a unique PSNE (three strict and two weak dominance solvable games) and the remaining four exhibited three PSNE.

The results are given in Table XXVIII, which identifies the results from each study grouped by the number and type of NE. Apart from the estimated Spearman correlation coefficients, the table also provides the number of elapsed generations and the scale factor that maximize the value of $\rho$. Overall, the Spearman correlations are high, the lowest value equal to 0.49 for games with a unique MSNE and the maximum value equal to 0.8857 for games with a unique PSNE. The NNs best described the average behavior of games with a unique PSNE, followed by games with a unique MSNE and finally games with three PSNE. The correlation coefficient of 0.55 for games with three PSNE is impressive given that these games create the biggest coordination problems and therefore are quite challenging to describe. In conclusion, the analysis is very encouraging regarding the similarity of behavior between experimental subjects and simulated NN agents.

## 7. CONCLUSION

This paper presented a single model for describing behavior and learning in the whole class of $n \times n$ randomly generated normal form games, regardless of how many PSNE the games exhibited or any other considerations. More importantly, this was achieved without reverting to a case-based approach and exogenously imposed measures of the similarity of games, as is usual done in the existing theoretical literature. The results of the neural network simulations in terms of how well they fit experimental human behavior were very encouraging. A high degree of correlation was found between the observed behavior of subjects in five different studies of one-shot games and the predicted behavior of the trained neural networks. The Spearman correlation coefficients ranged from 0.49 to 0.8857 which is particularly impressive given that the neural networks had never seen these games before and that these studies covered a wide variety of strategic games with different types and numbers of Nash equilibria.

In particular, there was an endogenous emergence in the NNs of behavior based on heuristics, such as $L1$ and $L2$, that have been found to explain humans' experimental behavior well. What is striking is that the neural networks were not predisposed or guided to learn these principles, they emerged from the simple property of myopic ex-post best response coupled with a backpropagation learning algorithm. The estimated coefficients of regressions performing rule extraction from the trained NNs showed significant differences when estimated separately for each type of game, as defined by the number of Nash equilibria in a game. The implication of this is that the NNs have indirectly learned to differentiate the games based on some similarity or classification measure that has arisen endogenously during the simulations.

The ability of the neural networks to learn principles of dominance and iterated dominance, with less success in the latter case also conforms to experimental results. Near convergence to the PSNE in games with a unique PSNE was observed in the simulation involving $2 \times 2$ games, with significant movement towards the PSNE in $3 \times 3$ games. A more definitive answer as to whether these NNs would converge in the $3 \times 3$ game simulations would require many more simulations as learning slows down appreciably in this case. The fact that the speed of learning slows down rapidly in $3 \times 3$ games is an important result as an agent may not be alive long enough to reach convergence and therefore this slowdown in the speed of learning may lead to a practical upper bound to agents' cognitive abilities. Such a problem might be especially important for humans as the number of games they will play in a lifetime or before they participate in an experiment is certainly constrained.

The NNs have also shown a preference towards playing risk dominant equilibria over payoff dominant equilibria, which is especially acute in the $2 \times 2$ game simulations, again in accordance with documented human behavior. In games with a unique mixed strategy Nash equilibrium, the neural networks' behavior was found to be positively correlated with the mixed strategy Nash equilibrium prescription although there were deviations as is also the case with human experimental data.

The learning literature in repeated coordination games commonly uses the principle of insufficient reason to assume uniform behavior or beliefs as the initial conditions to the learning algorithms they employ. However, equilibrium selection is dependent on initial play at time zero since the starting point can put play in basins of attraction of different Nash equilibria. One of the advantages of the trained NNs is that they can provide better estimates of the initial behavior of experimental subjects, which may lead to very different results regarding equilibrium selection because of path dependence.

Finally, an analysis of the relationship between payoffs and the relative sophistication of two players found that for games with less than three PSNE an increase in sophistication of an agent always leads to higher payoffs. Interestingly, as long as an agent remains more sophisticated than his opponent, increasing the sophistication of the opponent is beneficial, whereas in the opposite case it is detrimental. Hence, if sophisticated agents could choose which opponents to play against they would generally have an incentive to pick agents that are only slightly less sophisticated than themselves.

The possible directions for future research are many. Although this chapter has only explored the application of NNs for one-shot games, the above setup is sufficient to analyze repeated games as well. There is no restriction to providing the NNs with repeated games with exactly the same payoffs and analyzing their behavior, and if desired, including dependence on past behavior can be easily accomplished by adding lagged variables to the input layer.

Numerous other extensions can be implemented involving the setup of the population and the evolutionary processes governing it. This chapter has implicitly assumed the simultaneous birth of these agents into a population pool in which they participate equally. These agents however never died and no new births of agents occurred after the initial period. A more interesting case would be to allow for the birth and death of said agents at any time during the simulation. The birth and death rates could be a simple independent Markov chain process or could depend endogenously on other variables. For example, the probability of death in any generation could be a decreasing function of the payoffs the agent has received thus far. Allowing for this would make the networks even more realistic and would also probably allow the evolution or learning of a prospect theory type of utility curve. This would occur for example if survival is a step function of payoffs so that losses from negative payoffs would be more disadvantageous than gains of the same magnitude as they would bring agents closer to the survival cutoff value. Also, the birth of an agent into a population of experienced agents would also be worthwhile investigating. An interesting question would be whether this would help the agent to learn faster than entering a population of agents with no prior experience, or whether the more experienced networks would be able to consistently outsmart these younger agents on the basis of their experience.

Another avenue is to introduce social networking between agents. Network formation could be modeled by having agents which have accumulated more payoffs, or are wealthier, to interact more often with agents of greater wealth. This would cause groups to emerge whose pattern of play may be different and would also allow examinations of the emergence of inequality in the population. For example, the richer agents may become more risk seeking as they repeatedly encounter other rich agents for whom there is little risk of falling below the critical survival level. Another interesting extension is allowing agents to choose their opponents, for example they may seek to interact with the agents that have in the past history of play given the highest payoffs to a player e.g. agents that have cooperated instead of defected in past Prisoner's dilemma games.

The main goal of this paper was to determine whether NNs were appropriate models of generalized human learning, in contrast to another possible application which is to specifically design NNs that outperform other types of agents in simulations of this type. This is another subfield of the ACE literature that involves designing intelligent agents. Since maximizing the performance of the NNs was not part of the research agenda of this paper, a relatively simple NN was chosen to model agents, and a relatively simple standard training algorithm for the NNs was implemented. Future research could focus more on maximizing the performance of NNs when placed in such simulations or competitions. Other NN topologies could be considered instead of the multilayer feedforward network presented here. For example, networks with explicit memory capabilities, instead of the implicit memory capabilities of feedforward networks could be implemented. Other approaches could be the use of Support Vector Machines (SVM), or recurrent neural networks which would be particularly useful for repeated games. When designing such agents, important considerations are the performance of the agent, the speed of learning and its flexibility to adapt. This is of paramount importance if the agent is to be competing in an environment where other agents are also learning i.e. a non-stationary environment. In this respect replacing the backpropagation algorithm with more complex training algorithms, in particular ones that incorporate second-order information, could yield good results.

*Department of Economics*

*The University of Sydney*

*New South Wales, Australia*

lspi4871@usyd.edu.au

REFERENCES

Binmore, K., J. Swierzbinski, and C. Proulx (2001). Does Minimax Work? An Experimental Study. *Economic Journal 111*(473), 445–464.

Cabrales, A., W. Garcia-Fontes, and M. Motta (2000). Risk dominance selects the leader: An experimental analysis. *International Journal of Industrial Organization 18*, 137 – 162.

Cho, I. and T. Sargent (1996). Neural Networks for Encoding and Adapting in Dynamic Economies. *Handbook of Computational Economics 1*, 441–470.

Cooper, R. W., D. DeJong, R. Forsythe, and T. Ross (1994). *Problems of Coordination in Economic Activity*, Chapter Alternative Institutions for Resolving Coordination Problems: Experimental Evidence on Forward Induction and Preplay Communication. Kluwer Academic Publishers.

Cooper, R. W., D. V. DeJong, R. Forsythe, and T. W. Ross (1990). Selection criteria in coordination games: Some experimental results. *The American Economic Review 80*(1), 218–233.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica 69*(5), 1193–1236.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems 2*, 303–314.

Dror, I. E. and D. P. Gallogly (1999). Computational analyses in cognitive neuroscience: In defense of biological implausibility. *Psychonomic Bulletin & Review 6*(2), 173–182.

Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks 2*, 183–192.

Germano, F. (2007). Stochastic Evolution of Rules for Playing Finite Normal Form Games. *Theory and Decision 62*(4), 311–333.

Gilboa, I. and D. Schmeidler (1995). Case-Based Decision Theory. *The Quarterly Journal of Economics 110*(3), 605–639.

Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibirum Selection in Games*. Cambridge, MA: MIT Press.

Haruvy, E. and D. Stahl (2004). Deductive versus inductive equilibrium selection: experimental results. *Journal of Economic Behavior and Organization 53*(3), 319–331.

Hensher, D. and L. Johnson (1981). *Applied Discrete-choice Modelling*. Wiley.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks 4*, 251–257.

Hosmer, D. and S. Lemeshow (1989). *Applied Logistic Regression*. John Wiley & Sons New York.

Huysmans, J., B. Baesens, and J. Vanthienen (2006). Using rule extraction to improve the comprehensibility of predictive models. Technical Report KBI 0612, Katholieke Universiteit Leuven.

Ivanov, A. (2006, October). Strategic play and risk aversion in one-shot normal-form games: An experimental study.

Katz, K. (1996). *Three Applications of Game Theory*. Ph. D. thesis, University of Pennsylvania.

Kettner, R., J. Marcario, and N. Port (1993). A neural network model of cortical activity during reaching. *Journal of Cognitive Neuroscience 5*, 14–33.

Kuan, C.-M. and T. Liu (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics 10*(4), 347–364.

Leake, D. (Ed.) (1996). *Case-Based Reasoning: Experiences, Lessons, and Future Directions*. AAAI/MIT Press.

LeCun, Y., L. Bottou, G. B. Orr, and K.-R. Muller (1998). *Neural Networks: Tricks of the Trade*, Chapter Efficient backprop, pp. 9–50. Springer Berlin / Heidelberg.

Lehky, S. R. and T. J. Sejnowski (1988). Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature 333*, 452–454.

Lemke, C. E. and J. Howson, J. T. (1964). Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics 12*(2), 413–423.

Leung, M. T., H. Daouk, and A. S. Chen (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting 16*, 173–190.

LiCalzi, M. (1995). Fictitious play by cases. *Games and Economic Behavior 11*, 64–89.

Louviere, J., D. Hensher, and J. Swait (2000). *Stated choice methods*. Cambridge University Press New York.

Matlab (2007). *Mathworks, Inc., Natick, MA*.

Mazzoni, P., R. A. Andersen, and M. I. Jordan (1991). A more biologically plausible learning rule than backpropagation applied to a network model of cortical area 7a. *Cerebral Cortex 1*, 293–307.

McFadden, D., I. of Urban & Regional Development, and B. U. of California (1973). *Conditional Logit Analysis of Qualitative Choice Behavior*. Institute of Urban and Regional Development, University of California.

Mookherjee, D. and B. Sopher (1997). Learning and Decision Costs in Experimental Constant Sum Games. *Games and Economic Behavior 19*(1), 97–132.

Myers, R. (1986). *Classical and modern regression with applications*. Duxbury Press Boston, Mass.

Nagel, R. (1995). Unraveling in Guessing Games: An Experimental Study. *The American Economic Review 85*(5), 1313–1326.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters 86*(373-378).

Nguyen, D. and B. Widrow (1990). Improving the learning speed of 2–layer neural network by choosing initial values of the adaptive weights. In *IEEE First International Joint Conference on Neural Networks*, pp. 21–26.

Ockenfels, A. and R. Selten (2005). Impulse Balance Equilibrium and Feedback in First Price Auctions. *Games and Economic Behavior 51*, 155–170.

Pao, Y. (1989). *Adaptive pattern recognition and neural networks*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

Rey Biel, P. (2004). Equilibrium Play and Best Response to (Stated) Beliefs in Constant Sum Games. Technical report, mimeo.

Rieskamp, J., J. Busemeyer, and T. Laine (2003). How do people learn to allocate resources? Comparing two learning theories. *Journal of experimental psychology. Learning, memory, and cognition 29*(6), 1066–1081.

Robinson, T. (2000). Biologically plausible back-propagation. Technical report, Victoria University of Wellington.

Sargent, T. S. (1993). *Bounded Rationality in Macroeconomics*. Clarendon Press.

Sarkar, D. (1995). Methods to speed up error back-propagation learning algorithm. *ACM Computing Surveys 27*(4), 519–544.

Sarle, W. (1994). Neural networks and statistical models. *Proceedings of the Nineteenth Annual SAS Users Group International Conference*, 1538–1550.

Schotter, A., K. Weigelt, and C. Wilson (1994). A laboratory investigation of multiperson rationality and presentation effects. *Games and Economic Behavior 6*(3), 445–468.

Selten, R. (1998). Features of Experimentally Observed Bounded Rationality. *European Economic Review 42*(3-5), 413–36.

Selten, R., K. Abbink, and R. Cox (2005). Learning Direction Theory and the Winner's Curse. *Experimental Economics 8*(1), 5–20.

Sgroi, D. and D. J. Zizzo (2002). Strategy Learning in 3x3 Games by Neural Networks. Technical report, Department of Applied Economics, University of Cambridge.

Sgroi, D. and D. J. Zizzo (2007). Neural networks and bounded rationality. *Physica A: Statistical Mechanics and its Applications 375*(2), 717–725.

Smith, K. A. and J. N. Gupta (2002). *Neural networks in business : techniques and applications*. Idea Group Publishing.

Stahl, D. and E. Haruvy (2004). Rule Learning Across Dissimilar Symmetric Normal-Form Games.

Stahl, D. O. and P. W. Wilson (1995). On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior 10*(1), 218–254.

Steiner, J. and C. Stewart (2006). Learning by similarity in global games. http://www.econ.ed.ac.uk/papers/Learning_by_Similarity.pdf.

Straub, P. G. (1995). Risk dominance and coordination failures in static games. *The Quarterly Review of Economics and Finance 35*(4), 339–363.

Tang, F. F. (2001). Anticipatory learning in two-person games: some experimental results. *Journal of Economic Behavior & Organization 44*(2), 221–232.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life 8*(1), 55–82.

Tesfatsion, L. and K. L. Judd (2006). *Handbook of Computational Economics Volume 2*. Elsevier/North-Holland (Handbooks in Economics Series).

Walczak, S. and N. Cerpa (1999). Heuristic principles for the design of artificial neural networks. *Information and Software Technology 41*, 107–117.

Waldrop, M. (1992). *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster: New York.

Yang, Z. R., M. B. Platt, and H. D. Platt (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research 44*, 67–74.

Zipser, D. and R. A. Andersen (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature 331*, 679–684.

# A. TECHNICAL PRESENTATION OF NEURAL NETWORKS

## A.1. *Detailed formulation of feedforward neural networks*

Figure 6 is a detailed technical diagram of the topology of a feedforward neural network. The first layer is the input layer and each input neuron is denoted by $p_r$ where $r = 1,...,R$. Each $p_r$ is the payoff for a specific player from a specific cell for each game that will be played. For example, in a two player game with two actions for each player the value of R would be eight. The second layer consists of $S$ neurons, each of which is connected to all the input neurons, resulting in a total of $R \cdot S$ connections between the first and second layers. Each connection is associated with a weight, $w^{2,1}_{s,r}$, with $s,r$ denoting a connection from the $rth$ neuron to the $sth$ neuron and where the superscript $2,1$ represents that these weights are between the first and second layers of the NN. The activation of each neuron in the second layer, $i^2_s$, is the summation of the product of the inputs and their corresponding weights plus a constant or bias, $b^2_s$. For example, for each of $S$ neurons in the second or hidden layer:

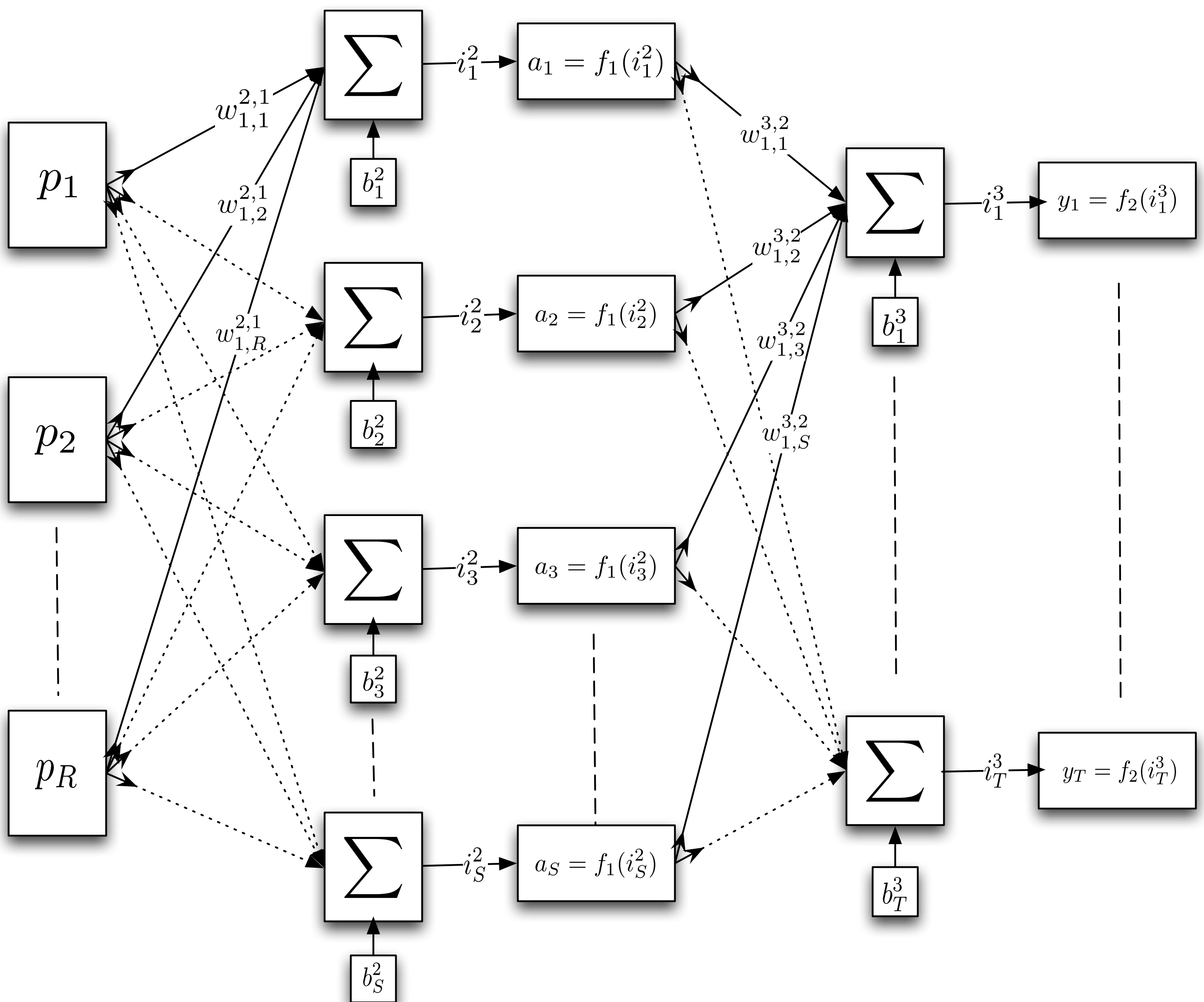
$$i^2_s = b^2_s + \sum_{r=1}^{R} w^{2,1}_{r,s} \cdot p_r \tag{2}$$

These inputs are now passed through a non-linear function, often called the squashing function, $f_1$, in this particular case the hyperbolic tangent sigmoid transfer (or tansig) function, $f_1(i_s) = 2 \cdot (1+e^{-2i^2_s})^{-1} - 1$ which maps values from $-\infty$ to $+\infty$ to the interval $(-1,1)$. The resulting outputs, $a_s$, are passed to the the final or output layer which is comprised of $T$ neurons. Each neuron will output the probability of each available action being played, hence $T$ is equal to the number of actions available to each player. Again each neuron in the output layer is connected to every neuron in the second layer with connection weights, $w^{3,2}_{s,t}$. The input to each $t$ neuron is the summation of product of the outputs, $a_s$, and the corresponding weights, $w^{3,2}_{s,t}$ plus a bias $b^3_t$:

$$i^3_t = b^3_t + \sum_{s=1}^{S} w^{3,2}_{s,t} \cdot a_s \tag{3}$$

These inputs are transformed by the function, $f_2$, resulting in the final outputs of the NN, $y_t$. In order for these outputs to be interpretable as probabilities of playing each action they must sum to one i.e. $\sum_{t=1}^{T} y_t = 1$. This is achieved by using a softmax function where $y_t = f_2(i^3_t) = \frac{e^{i^3_t}}{\sum_{t=1}^{T} e^{i^3_t}}$. In the case of two action games, the number of output neurons, $T$, is equal to two, so that the resulting values of these two neurons are the probabilities of playing each action.

FIGURE 6.— Detailed structure and topology of feedforward neural networks



Knowledge is stored in NNs by the weights and biases of all the neurons, which is why it is referred to as distributed knowledge since it is not localized in any specific region of the NN structure. Hence, learning in a NN is accomplished through the updating of the weights and biases after presentation of each set of inputs, in this case each game's payoff matrix. In supervised learning, for each set of inputs, $P = \{p_1, ..., p_R\}$, there exists a set of ideal outputs, $Z = \{z_1, ..., z_T\}$.

The question now arises as to what the agent shall regard as the ideal or correct output, since in contrast to Sgroi and Zizzo (2002) there is no external teacher to provide this. Selten (1998), one of the key proponents of bounded rather than perfect rationality of agents, puts forth a qualitative theory called learning direction theory. Although not a fully fledged theory it is more of a general qualitative principle of learning from which more specific quantitative learning models can be fashioned. Selten (1998) argues that a general conclusion that can be drawn from the experimental literature is that the general principle guiding learning is ex-post rationality. This implies that an economic agent will move or modify his action in the direction of ex-post best response to the immediately prior outcome. Implementations of learning direction theory include an application to the Winner's Curse in Selten et al. (2005), an explanation of how people learn to allocate resources in Rieskamp et al. (2003), modeling behavior in guessing games Nagel (1995) and an application to auction theory in Ockenfels and Selten (2005).

In the spirit of ex-post rationality the ideal output will be defined as the hypothetical best response of the NN after observing the action chosen by its opponent. Hence, exactly one $z_t$ will be equal to one and the rest will be equal to zero for each set of inputs. Define the mean square error, $E$, of the network, to be:

(4) $$E = \frac{1}{2} \cdot \sum_{t=1}^{T} (z_t - y_t)^2$$

The most common learning algorithm used in training NNs is the backpropagation algorithm which uses a gradient descent technique. After the presentation of each set of inputs the weights are changed according to the following equation:

(5) $$\triangle w^{\cdot,\cdot}_{\cdot,\cdot} = -\eta \frac{\partial E}{\partial w^{\cdot,\cdot}_{\cdot,\cdot}}$$

This is a gradient descent technique as the updating of the weights depends on the negative of the gradient of the error function and on its magnitude, where $\eta$ is a constant referred to as the step size (or learning rate) that governs the magnitude of the change in the weights. Hence, weights will be changed in the direction which reduces the error, $E$, and the magnitude of the change will also be related to the sensitivity of the error function to small changes in the weight. The necessary algebra to derive $\partial E/\partial w$ for both output layer and hidden layer neurons is presented below.

In more detail, for weights in the output layer, using the chain rule leads to the following derivation:

$$\frac{\partial E}{\partial w^{3,2}_{s,t}} = \frac{\partial E}{\partial y_t}\frac{\partial y_t}{\partial i^3_t}\frac{\partial i^3_t}{\partial w^{3,2}_{s,t}} \tag{6}$$

However, from equation 3 it is clear that:

$$\frac{\partial i^3_t}{\partial w^{3,2}_{s,t}} = a_s \tag{7}$$

and from equation 4:

$$\frac{\partial E}{\partial y_t} = (y_t - z_t) \tag{8}$$

Substituting these equations into equation 6 results in:

$$\frac{\partial E}{\partial w^{3,2}_{s,t}} = (y_t - z_t)a_s f_2^{'}(i^3_t) \tag{9}$$

The necessary calculations for weights in hidden layers is more involved as the desired output of such neurons is not immediately available as is the case for output layer neurons. Using the chain rule, the analog to equation 6 for a hidden layer neuron is:

$$\frac{\partial E}{\partial w^{2,1}_{s,r}} = \sum_{t=1}^{T}\frac{\partial E}{\partial y_t}\frac{\partial y_t}{\partial a_s}\frac{\partial a_s}{\partial w^{2,1}_{s,r}} \tag{10}$$

This equation now has a summation of terms over $t$ since hidden layer weights can affect the error of the NN through all the output layer neurons due to the propagation of the effect of $w^{2,1}_{s,r}$ through the interconnections between the $sth$ neuron and all $T$ neurons in the output layer. The derivative of the output of the $sth$ neuron with respect to the weight under investigation is given by:

$$(11) \qquad \frac{\partial a_s^2}{\partial w_{s,r}^{2,1}} = p_r f_1^{'}(i_s^2)$$

The derivative of the error function w.r.t. the output of each final layer neuron, $\partial E / \partial y_t$, is still given by equation 8. Finally, the derivatives of the output of each final layer neuron w.r.t. the output of each hidden layer neuron are given by:

$$(12) \qquad \frac{\partial y_t}{\partial w_{s,r}^{2,1}} = w_{t,s}^{3,2} f_2^{'}(i_t^3)$$

In conclusion, substituting equations 11, 8 and 12 into equation 10 leads to the following equation, which is well defined as both $f_1$ and $f_2$ are differentiable functions:

$$(13) \qquad \frac{\partial E}{\partial w_{s,r}^{2,1}} = p_r f_1^{'}(i_t^2) \sum_{t=1}^{T} (y_t - z_t) w_{t,s}^{3,2} f_2^{'}(i_t^3)$$

## B. ROBUSTNESS OF SIMULATIONS

The behavior of the agents in the simulations is a function of the number of layers and neurons, of the initial starting weights and of the learning rule (including the learning parameter rate). The number of layers and neurons are hypothesized to influence the behavior of the networks especially through their effects on the complexity and bounded rationality of the networks.

However, the implicit assumption is that differences in initial weights and learning rates (within reasonable values) have no significant effect on NN learning. The reference to "reasonable" learning rates is made because with extreme learning rates (near zero or one) the NN will not learn at all. However, there is a range of suitable parameter values where learning is viable and for this range the end results should be very similar. Since there is no a priori reason to select specific initial weights or learning rates it is necessary to check for the robustness of results with respect to variations in these parameters.

A common observation in standard uses of NNs such as forecasting is that the effect of the initial weights is significant and may not necessarily wear off with more learning i.e. there exists a form of history dependence. In particular, because the learning algorithms employed in NNs are susceptible to converging on a local minimum rather than a global minimum certain initial values for weights may lead to significantly worse behavior. However, there is an important difference between the usual forecasting applications of NNs and the application employed in this paper. In standard forecasting applications, the NN is presented with a fixed input and desired output vector so that if a learning algorithm becomes stuck at a local minimum it will remain there. However, in this application neither the input nor the desired output are fixed. The former varies because of the random sampling of each game's payoffs and the latter because the decision rule of an agent's opponent is stochastic. Hence, even if the backpropagation learning procedure happens to be led to an undesirable region of the error surface, given enough generations the weights will always escape this region due to the stochastic nature of the system. Hence, in this application initial weights may affect the speed of convergence to a steady state, but given a large enough number of rounds they should not affect convergence to the global minimum.

### B.1. *Robustness to initial weights*

The initial weights for the NNs throughout this paper are determined using the Nguyen and Widrow (1990) technique. The dependence of the trained network responses on initial weights will be tested by running ten simulations where the networks are randomly initialized with different weights and examining the correlation of the respective network outputs after training. All ten simulations were run for 100,000 generations with exactly the same training data sets and the resulting trained NNs, all of topology $\psi\{50,3\}$, were then presented with the same test data set. Finally, the outputs of all the networks in a simulation were averaged to give the representative behavior of that particular simulation. It is these averaged outputs that the Spearman correlation coefficients were calculated on. In every possible pairing of simulations, the Spearman correlation coefficient was always greater than 0.99, a very strong indication that initial weights do not significantly influence the long-run learning of a population of neural networks.

Another measure of the dependence of NN behavior on initial weights is to calculate the Mean Absolute Difference (MAD) between average simulation outputs. The network averaged MAD, calculated on all possible combinations of simulations, was found to be only 0.0222. This is extremely strong evidence that the simulations do not exhibit long run dependence on the initial starting parameters of the NNs.

### B.2. *Robustness to learning rate parameters*

The learning rate often utilized in the literature for online learning is equal to a value of 0.01, although this is not a steadfast rule as it depends on the application. Robustness of the learning results of trained NNs will be examined by varying the learning rate from a minimum value of 0.0025 to a maximum value of 0.05, and then analyzing the similarity in NN outputs across the learning rates. This will be accomplished by computing the pairwise Spearman correlation coefficients between the outputs of the trained NNs. Once again, the Spearman correlation coefficient was always found to be greater than 0.99 implying that the value of the learning rate parameter was not an influential factor on NN learning.

The network averaged MAD for this case is 0.0324, which again is very reassuring concerning the limited impact of the learning rate on the learned behavior of the NNs. This also implies that the impact of changing the learning rate on the average speed of agents' learning is minimal. These results are reassuring as to the robustness of the approach employed in this paper. NN practitioners may be surprised by such strong robustness, however it is a consequence of the subtleties of this application, namely the stochastic nature of the learning process due to the online learning procedure, the random sampling of games and the stochastic decision rule employed.

### B.3. *Experimentation with other more sophisticated algorithms*

The main results of this paper are based on the use of the backpropagation algorithm on the pretence that it is more biologically reasonable than other more complex algorithms used in the neural networks field. Preliminary investigation of simulations with other algorithms do not reveal a dependence of the results on the learning algorithm used. The driving factor seems to be the basic assumption of learning by adapting internal weights according to myopic best response. The exact mechanics and calculations of weight adjustments in order to better respond do not seem to have any significant effect apart from changing the speed of learning (and even the difference in speed of learning was not particularly acute). Also, the learning process and the types of heuristics discovered by the neural networks still remain the same.

The results of trial runs from other techniques such as backpropagation with momentum, backpropagation with variable step size (Sarkar, 1995), or very sophisticated algorithms that employ second-order techniques, such as conjugate gradient methods, support the conjecture that the driving force of the results of the simulations is the myopic best response assumption, instead of the exact method of updating the NNs' weights.

## C. DETAILS ON THE COMPUTATION OF GAME PROPERTIES

### C.1. *Computing the mixed strategy equilibria of* $3 \times 3$ *games*

Contrary to the trivial calculations involved in computing the mixed strategy of a $2 \times 2$ game, namely solving a system of two simultaneous equations in two variables, the calculation of such equilibria in games with more than two actions is more involved. The main complication arises because in two action games it is necessarily true that a mixed strategy equilibrium will entail non-zero probabilities for both actions i.e. there exists an interior solution. When there exist more than two actions, then it is possible for a mixed strategy equilibrium to include corner solutions where some actions may not be played at all. For example, in games with three actions, only one action may have a zero probability of play, as the support of a MSNE requires at least two actions.

Let $p,q$ denote mixed strategies for the row and column player respectively with $m$ and $n$ actions each. Then $p \in \Delta_m = \{p \in \mathbb{R}^m : p \geq 0, p^T e = 1\}$ , and likewise $q \in \Delta_n = \{q \in \mathbb{R}^n : q \geq 0, q^T e = 1\}$, where $e$ is a column vector whose dimension is dictated by the context. Let the payoff matrices of the game be $A,B \in \mathbb{R}^{m\times n}$ where $A$ are the payoffs to the row player and $B$ for the column player. Hence, the expected payoffs to the row and column players are $p^T Aq$ and $p^T Bq$ respectively. A mixed strategy equilibrium $(p_*, q_*)$ is given by:

$$
\begin{aligned}
p_*^T A q_* &\geq p^T A q_* \\
p_*^T B q_* &\geq p_*^T B q \\
&\forall p,q
\end{aligned}
\tag{14}
$$

The trick to solving this is to reformulate this problem in an equivalent manner that can be solved as a linear complementarity problem. Lemke and Howson (1964) showed that the above problem has a dual linear complementarity problem and also provided an algorithm to solve it. Namely, they show that equation 14 can be represented by the following optimization problem. Let $e$ be a column vector and $E$ a square matrix whose dimensions are determined by the context.

$$
\begin{aligned}
k \in \mathbb{R} : (kE - B^T) &> 0 \\
k \in \mathbb{R} : (kE - A) &> 0 \\
(kE - B^T)q &\geq e \\
(kE - A)q &\geq e \\
q^T[(kE - B^T)p - e] &= 0 \\
p^T[(kE - A)q - e] &= 0 \\
p,q &\geq 0 \\
p_* &= \frac{p}{p^T e} \\
q_* &= \frac{q}{q^T e}
\end{aligned}
\tag{15}
$$

Since all games have payoffs bounded between -1 and 1, setting $k = 1$ is sufficient to comply with the first two inequalities. Lemke and Howson (1964) provide an algorithm to solve this type of optimization problem. This algorithm was implemented to calculate the MSNE for games which had no PSNE.

TABLE XXIX

GENERALIZED COORDINATION GAME

| | | Player 2 | |
|---|---|---|---|
| | | Left | Right |
| Player 1 | Up | A,a | B,b |
| | Down | C,c | D,d |

## C.2. *Computing the risk dominant equilibria in games*

A significant topic for coordination games is equilibrium selection as Nash equilibrium theory is silent about which equilibrium will be played. Many theoretical refinements have been proposed such as subgame perfection or trembling hand. Harsanyi and Selten (1988) propose a solution concept, termed risk dominance, where players are more likely to play the equilibrium which poses the least risk for the two players. In $2 \times 2$ games, they define a risk dominant equilibrium by comparing the Nash products of equilibria, which are simply the product of the losses arising for each player from unilateral deviation from the equilibrium. For example, in a generalized coordination game as in Table XXIX (Up, Left) would risk dominate (Down, Right) if $(A-C)(a-b) > (D-B)(d-c)$.

Although the risk dominance concept yields a simple method of determining the risk dominant equilibrium for $2 \times 2$ games, this is not the case for games with a larger action space. The following procedure taken from Haruvy and Stahl (2004) for determining the risk dominant equilibrium in $3 \times 3$ games, encompasses the simple Nash product solution for $2 \times 2$ games. They justify a Bayesian approach where each player's second-order beliefs (i.e. beliefs about what an opponent's beliefs are) are uniformly distributed. For example, for $n$ actions, player 1 would expect player 2's beliefs about his own play to be uniformly distributed on an $n$-dimensional simplex, likewise for player 2.

1. For a game with $n$-actions, let $A^{NE}$ be the set of actions which can lead to a Nash equilibrium and let $\nabla(A^{NE})$ be the simplex on $A^{NE}$.
2. For each player, $i$, and each action in $A^{NE}$, $j$, calculate $p_i^j(rd)$ as the proportion of $\nabla(A^{NE})$ where action $j$ is a best response given the assumption that an opponent's beliefs are uniformly distributed.
3. Finally, the risk dominant action for player $-i$ will be the action $-j$ which maximizes expected payoffs assuming that the opponent plays according to $p_i^j(rd)$ for all $j$.

Given that all probabilities must sum to one, this problem can be expressed in terms of the probability triplet *(x,y,1-x-y)*. Hence, step 2 amounts to the triple integration:

$$p_i^j(rd) = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} I dz dy dx \tag{16}$$

$$I = \begin{cases} 1 & \textit{if action j is player i's best response to an} \\ & \textit{opponent's mixed strategy (x,y,1-x-y)} \\ 0 & \textit{otherwise} \end{cases}$$

The previous attempts to explain coordination failures hinged on deductive principles, however another strain of research exists that relies on inductive principles. In particular, equilibrium selection is seen as a result of specific dynamic learning processes employed by players. An experimental comparison of these two strands of theoretical research is provided by Haruvy and Stahl (2004) who find that an inductive explanation performs better than deductive explanations, with the latter failing miserably in some games. The drawback of inductive explanations is that they may not always provide an answer as they do not necessarily converge. The approach of this paper is necessarily inductive as equilibria will be selected as a result of learning processes.

## D. FURTHER ANALYSES OF $\Psi^2_{HET}$ AND $\Psi^2_{HOM}$ SIMULATIONS

### D.1. *An alternative technique of examining convergence to a steady state in* $\Psi^2_{hom}$

Verification that agent learning has settled down to a steady state result can also be determined by examining whether the distribution of network responses has settled down. Figure 7 graphs the evolution of the distribution of NN outputs of $2 \times 2$ games in $\Psi^2_{hom}$ i.e. the probability of playing an action, throughout training. The probabilities are placed into ten bins so that the y-axis is the percentage of responses which fall into each particular bin[23]. It is reasonable to expect that NN behavior will differ markedly depending on the number of NE of the training set games. Hence, three separate graphs are provided the first for games with only a MSNE, the second for games with a unique PSNE, and the third for games with two PSNE and one MSNE. The starting observations should be approximately uniformly distributed with subsequent learning altering this.

For games with a single MSNE, the probability distribution of network responses appears to stabilize much earlier than the end of the training run as is exhibited by the constant height of the shaded regions. As the NNs learn they are increasingly avoiding playing actions with extreme probabilities close to zero or one and tend to give responses around the midpoint probability of 0.5. This is reasonable as by definition MSNE require agents to mix amongst actions rather than play a single action with probability one.

[23] In the case of equiprobable play, the height of each bin would be the same.

As regards games with a unique PSNE the probability distribution of the networks responses settles down quite quickly to a stable distribution where almost all responses are very close to zero or one, corresponding to the networks playing an action with near certainty.

In games with two PSNE the results are more complex. There is an initial phase, roughly the first 100,000 generations, where the probability of extreme responses increases with a corresponding decrease in probabilities of returning less extreme decisions. Thereafter there seems to be a roughly constant probability of responses in all bins except the extreme ones. However, even towards the end of the simulations there seems to be variability in the probabilities of extreme responses which however can be explained by the coordination problems exhibited by games with two PSNE. If by chance the networks start coordinating on one of the two PSNE equilibria then their responses will increase in favor of one action at the expense of the other. However the stochastic nature of decisions, or actions actually taken, means that there will always be tension between the two PSNE and therefore there will be oscillations between the two caused by the stochastic nature of the decisions rules and the random sampling of game payoffs.

### D.2. *Payoff analysis of NNs in the* $\Psi^2_{het}$ *simulation*

Research into bounded rationality has shown that more sophisticated players need not necessarily perform better when pitted against less sophisticated players. A natural question to ask is whether NNs with more complex topologies outperform simpler NNs. Table XXX breaks down the payoffs of each NN topology/agent according to the number of PSNE in each game. All network topologies yield an expected payoff of zero for games with unique MSNE, but payoffs are significantly different from zero for games with one and two PSNE. As far as games with a unique PSNE are concerned the linear network, $\psi_1\{\varnothing,\varnothing\}$, performs much worse than the rest of the non-linear topologies. When it comes to games with two PSNE however there is no clear relationship between sophistication and performance. There exist three non-linear topologies, $\psi_6\{20,2\}$, $\psi_8\{50,1\}$ and $\psi_9\{50,2\}$ that have amassed less payoffs than the linear network, however the best documented performance of 0.1471 for $\psi_4\{5,3\}$ is much higher than that of $\psi_1\{\varnothing,\varnothing\}$, 0.1066.

### D.3. *Further analysis of* $\Psi^2_{het}$ *simulation*

As networks in this simulation are less sophisticated in complexity compared to $\Psi^2_{hom}$, it is reasonable to assume that they will be more likely to err and play a dominated strategy, or not perform iterated dominance calculations. This is confirmed, as the average probability of playing a dominant action is equal to 0.9644 and the probability of performing iterated dominance is equal to 0.9147, both of which are less than in $\Psi^2_{hom}$, but are still quite impressive. Breaking down these results into types will lead to additional information about

FIGURE 7.— Distribution of NNs' output during training in $\Psi^2_{hom}$

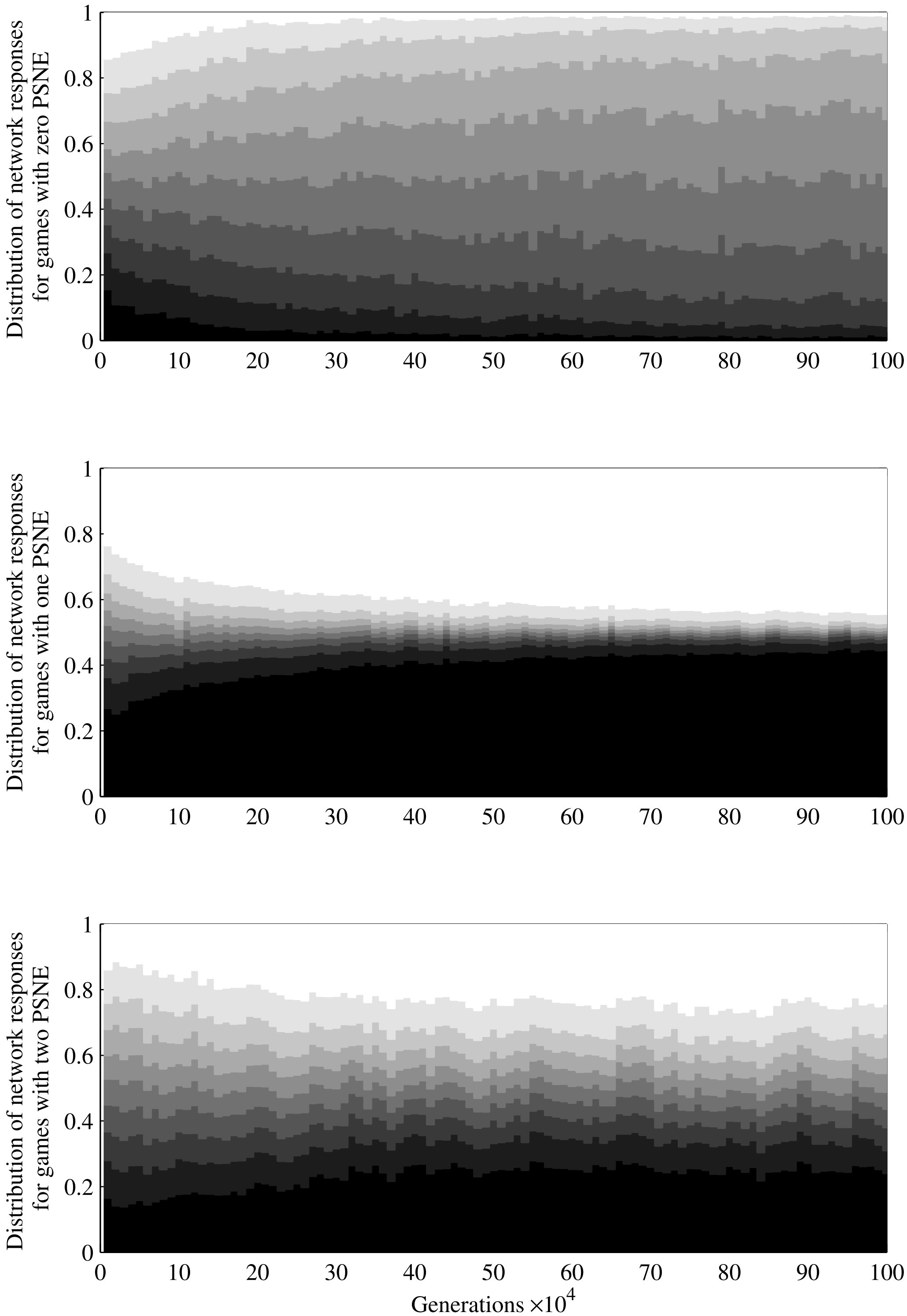


TABLE XXX

INDIVIDUAL NN MEAN PAYOFFS IN $\Psi^2_{het}$

| NN agents | No. of PSNE 0 | 1 | 2 | Any |
|---|---|---|---|---|
| $\psi_1\{\varnothing,\varnothing\}$ | 0.0506 | 0.2122 | 0.1066 | 0.1782 |
| $\psi_2\{5,1\}$ | -0.0086 | 0.2859 | 0.1099 | 0.227 |
| $\psi_3\{5,2\}$ | 0.0049 | 0.2949 | 0.1315 | 0.2378 |
| $\psi_4\{5,3\}$ | 0.0097 | 0.3037 | 0.1471 | 0.2462 |
| $\psi_5\{20,1\}$ | 0.0126 | 0.2997 | 0.1214 | 0.2413 |
| $\psi_6\{20,2\}$ | -0.0232 | 0.3166 | 0.0887 | 0.2472 |
| $\psi_7\{20,3\}$ | -0.0136 | 0.3262 | 0.1093 | 0.2564 |
| $\psi_8\{50,1\}$ | -0.0005 | 0.302 | 0.1031 | 0.2386 |
| $\psi_9\{50,2\}$ | 0.0005 | 0.3185 | 0.0923 | 0.2503 |
| $\psi_{10}\{50,3\}$ | -0.006 | 0.3148 | 0.1295 | 0.2514 |

TABLE XXXI

PROBABILITY OF NN COMPLIANCE WITH DOMINANCE AND ITERATED DOMINANCE PRINCIPLES IN $\Psi^2_{het}$

| NN agent | Probability dominant play | Probability best responding to opponent |
|---|---|---|
| $\psi_1\{\varnothing,\varnothing\}$ | 0.8461 | 0.6797 |
| $\psi_2\{5,1\}$ | 0.9533 | 0.8687 |
| $\psi_3\{5,2\}$ | 0.9653 | 0.919 |
| $\psi_4\{5,3\}$ | 0.9699 | 0.9254 |
| $\psi_5\{20,1\}$ | 0.982 | 0.9559 |
| $\psi_6\{20,2\}$ | 0.9911 | 0.9587 |
| $\psi_7\{20,3\}$ | 0.9892 | 0.9558 |
| $\psi_8\{50,1\}$ | 0.9796 | 0.9519 |
| $\psi_9\{50,2\}$ | 0.9864 | 0.9642 |
| $\psi_{10}\{50,3\}$ | 0.9856 | 0.964 |

how performance is related to the number of neurons and layers, especially for the case of the linear network with no hidden layers. Table XXXI logs the probability that each network will adhere to the concepts of dominance and iterated dominance. The results for the linear NNs are strikingly different from the rest as they tend to violate dominance principles much more often than more complex players. There is a clear positive relationship between complexity and adherence of the networks to the principles of dominance and iterated dominance, although the largest increase occurs moving from the linear model to the simplest single hidden layer NN, $\psi_2\{5,1\}$.

Turning now to further analysis of behavior in games with a unique MSNE, a breakdown of the correlation between the MSNE prescription and NN output by type of network is interesting as it relates the level of bounded rationality to MSNE behavior. Table XXXII provides the Spearman correlation for each of the ten different network topologies involved in this simulation. The lowest correlation, essentially zero, is attributed to $\psi_1\{\varnothing,\varnothing\}$ probably because it is the most bounded network, only capable of linear associations between its inputs and outputs. The result from the $\Psi^2_{hom}$ simulation that the $\psi_1\{50,3\}$ network exhibited low correlation

TABLE XXXII
CORRELATION OF INDIVIDUAL NNS' OUTPUT AND THE MSNE PRESCRIPTION IN $\Psi^2_{het}$

| | NN agents | | | | |
|---|---|---|---|---|---|
| Correlation | $\psi_1\{\varnothing,\varnothing\}$ | $\psi_2\{5,1\}$ | $\psi_3\{5,2\}$ | $\psi_4\{5,3\}$ | $\psi_5\{20,1\}$ |
| $\rho_s$ | -0.0023 | 0.6548 | 0.4869 | 0.4747 | 0.6252 |
| | NN agents | | | | |
| Correlation | $\psi_6\{20,2\}$ | $\psi_7\{20,3\}$ | $\psi_8\{50,1\}$ | $\psi_9\{50,2\}$ | $\psi_{10}\{50,3\}$ |
| $\rho_s$ | 0.4629 | 0.2799 | 0.4938 | 0.3415 | 0.2508 |

between actual behavior and the MSNE prescription, is also found in this simulation for the one network with the $\psi\{50,3\}$ topology, where the Spearman correlation is only 0.25. A surprising result is that, excluding the linear network, the correlation with the MSNE prescription decreases as the number of layers and/or number of neurons increases.

### D.4. *Further analysis of MSNE behavior in* $\Psi^2_{hom}$

Examining the evolution of the correlation coefficient during training (calculated per 50,000 generations) reveals interesting behavior. Although not evident from Figure 8, the correlation coefficient starts off at approximately zero as outputs are initially random and quickly increases to a maximum of roughly 0.56 between 50,000 and 100,000 generations. From this point onwards the correlation coefficient begins falling as the number of generations increases. The lowest point occurs at the end of the training set where the correlation is fluctuating between 0.15 and 0.2.

Figure 9 plots three dimensions, the x- and y-axes are broken down into 100 bins each and the color of the graph denotes how many observations fell into these combinations of bins[24]. It is clear from the graph that the relationship is non-linear as the relationship is flat for MSNE values of between roughly 0.2 and 0.8, but at the endpoints there is a positive relationship from which it can be concluded that the networks are affected by the MSNE prescription only when they are close to zero or one. Snapshots of this density plot during the training phase reveal that initially there is a roughly linear relationship over all values of the MSNE probability which however changes over time and is finally described by Figure 9 at the end of the training period.

[24]If there were a perfect positive linear relationship we would expect to see a white diagonal line surrounded by a black area.

FIGURE 8.— Spearman rank correlation over generations for all networks in $\Psi^2_{hom}$

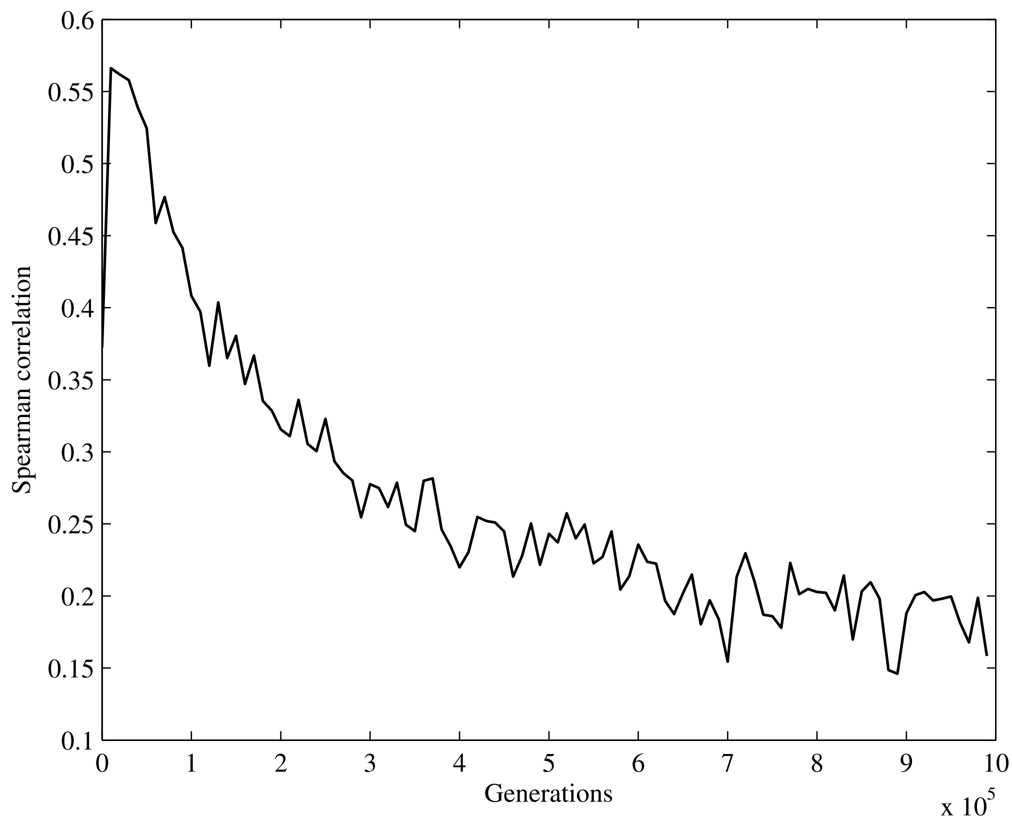
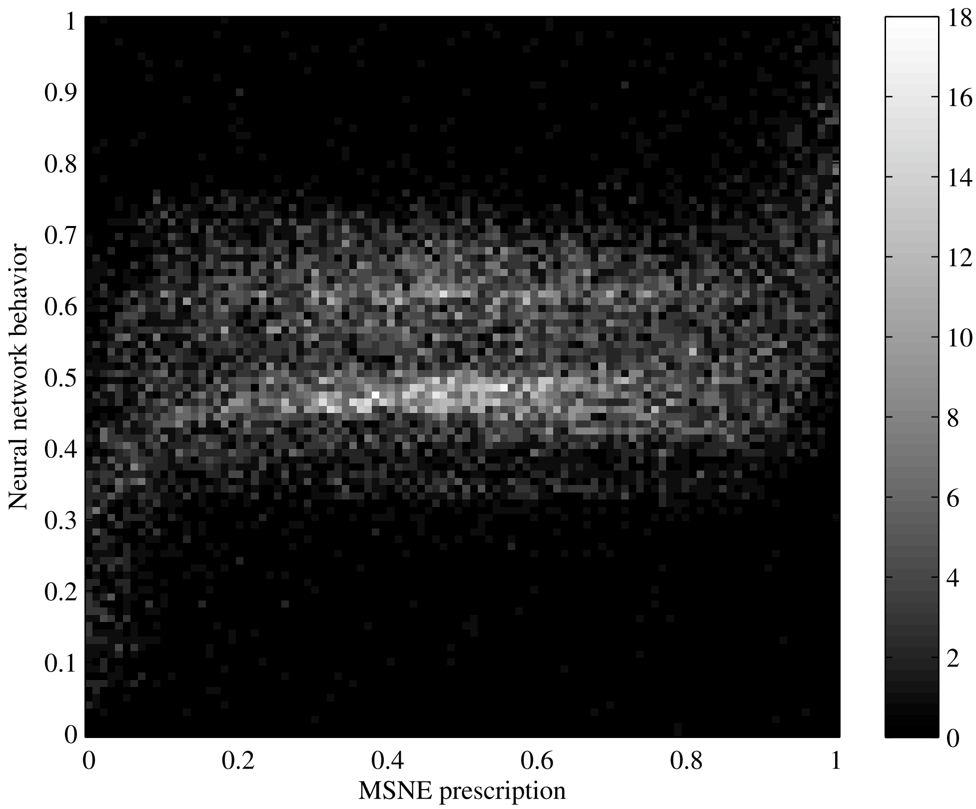


FIGURE 9.— Density plot of MSNE and actual network probabilities for $2 \times 2$ games




The reason why networks appear to be learning the MSNE in the first 100,000 generations and subsequently seem to change behavior so that it actually moves away from MSNE play is puzzling. One possible explanation is that because games with unique mixed strategy equilibria are under-represented in the training data set, games with unique PSNE may be interfering with learning the MSNE. This hypothesis is rejected however as balancing the training data does not affect the correlation significantly as can be seen in Table XXXVIII. Other possible hypotheses include effects that arise because of the number of agents. This indeed seems to play a role as the correlation for the $\Psi^2_{hom}$ simulation is 0.1821 which however increases to 0.3106 if the same simulation is performed with only two players in the population.

## E. RESULTS OF OTHER SIMULATIONS

This section discusses the main results of variations of the simulations discussed in the main text. Such variations in the simulations include using only simple linear NNs, changing the size of the population and changing the game payoff sampling mechanism.

Regarding the game payoff sampling mechanism, if the payoffs of games are randomly sampled from a bounded uniform distribution, then there is no control over the frequency of types of games appearing from such a sampling scheme. For example, it is easy to show that the sampling distribution of $2 \times 2$ games with zero, one and two PSNE is $\frac{1}{8}$, $\frac{3}{4}$ and $\frac{1}{8}$ respectively. During the training of the neural networks the sample proportion of different types of games may affect the speed and degree of learning. Also, it is possible that a high occurrence of one type of game may interfere with the learning of other types of games since more adjustments in the neural network parameters will be made for the most common type. Ideally, the proportion of games in the training dataset should be equal to that found in the environment so as to maximize the ecological validity of the simulation, however in practice this is difficult to ascertain. A robustness check by varying the sampling scheme can be performed to examine whether behavior is significantly affected by this. A balanced dataset shall be created where each type of game, as delineated by the number of PSNE, shall be sampled equally often. Results of simulations on this balanced dataset are provided in the Appendices.

To differentiate between the different simulations now that additional parameters are being varied it is necessary to amend the symbolism, $\Psi$, of the NN simulations. To accommodate for the possibility that the population may be comprised of only two NNs instead of ten NNs, a second superscript will be added denoting the number of agents. Also, to accommodate for the use of a balanced dataset instead of a uniform random sampling scheme, the letter $b$ will also be added to the superscript for the former and the letter $r$ for the latter. Finally, to symbolize the sole use of linear neural networks without any hidden layers a second subscript will be used $\varnothing$, referring to the fact that there are no hidden layers. For example, $\Psi^{3,2,b}_{hom,\varnothing}$ refers to a neural network trained on $3 \times 3$ games and a balanced dataset, comprised of two homogeneous NNs of topology $\psi\{\varnothing, \varnothing\}$.

Detailed results and comparisons of the following list of simulations are presented below.

$$\begin{aligned}
\Psi^{2,2,r}_{hom} &= [\psi_1\{50,3\}, \psi_2\{50,3\}] \\
\Psi^{2,10,b}_{hom,\varnothing} &= [\psi_1\{\varnothing,\varnothing\}, ..., \psi_{10}\{\varnothing,\varnothing\}] \\
\Psi^{2,2,b}_{hom,\varnothing} &= [\psi_1\{\varnothing,\varnothing\}, \psi_2\{\varnothing,\varnothing\}] \\
\Psi^{2,10,b}_{hom} &= [\psi_1\{50,3\}, ..., \psi_{10}\{50,3\}]
\end{aligned}$$

Various comparisons of interest can be made using these simulations. Firstly, $\Psi^{2,10,b}_{hom}$ can be compared to $\Psi^{2,10,r}_{hom}$ in an effort to examine whether balancing the training data set leads to significant changes in NN behavior. The two simulations incorporating only linear NNs can be used to ascertain whether they would be good candidates for modeling human behavior or whether more sophisticated NNs are needed. Also, the two simulations, $\Psi^{2,10,b}_{hom,\varnothing}$ and $\Psi^{2,2,b}_{hom,\varnothing}$ differ only in the number of agents in the simulation, in the second case there are only two agents allowing the examination of whether conventions can arise more easily when there are only two agents playing repeatedly against each other instead of ten randomly rematched agents. The same comparison can also be performed with more sophisticated NN agents by comparing simulations $\Psi^{2,2,r}_{hom}$ and $\Psi^{2,10,r}_{hom}$. The following tables will be referred to in the coming sections.

TABLE XXXIII

COMPARISON OF COMPLIANCE WITH DOMINANCE AND ITERATED DOMINANCE PRINCIPLES

| | Simulations | | | | |
|---|---|---|---|---|---|
| | $\Psi^{2,10,r}_{hom}$ | $\Psi^{2,2,r}_{hom}$ | $\Psi^{2,10,b}_{hom,\varnothing}$ | $\Psi^{2,2,b}_{hom,\varnothing}$ | $\Psi^{2,10,b}_{hom}$ |
| dominance | 0.988 | 0.9713 | 0.7203 | 0.7195 | 0.9748 |
| iterated dominance | 0.9736 | 0.9331 | 0.6092 | 0.6099 | 0.951 |

TABLE XXXIV

COMPARISON OF PROBABILITY OF JOINT NE PLAY ACCORDING TO NUMBER OF PSNE

| | Simulations | | | | |
|---|---|---|---|---|---|
| No. of PSNE | $\Psi^{2,10,r}_{hom}$ | $\Psi^{2,2,r}_{hom}$ | $\Psi^{2,10,b}_{hom,\varnothing}$ | $\Psi^{2,2,b}_{hom,\varnothing}$ | $\Psi^{2,10,b}_{hom}$ |
| 1 | 0.9585 | 0.8949 | 0.4079 | 0.4106 | 0.91.94 |
| 2 | 0.7519 | 0.7728 | 0.4985 | 0.5045 | 0.7779 |

TABLE XXXV

COMPARISON OF THE PROBABILITY OF JOINT PSNE PLAY ACCORDING TO NO. OF PLAYERS WITH DOMINANT ACTIONS

| | Simulations | | | | |
|---|---|---|---|---|---|
| No. of players with a dominant action | $\Psi^{2,10,r}_{hom}$ | $\Psi^{2,2,r}_{hom}$ | $\Psi^{2,10,b}_{hom,\varnothing}$ | $\Psi^{2,2,b}_{hom,\varnothing}$ | $\Psi^{2,10,b}_{hom}$ |
| 1 | 95.01% | 87.42% | 35.75% | 35.03% | 89.94% |
| 2 | 97.54% | 93.77% | 50.71% | 52.81% | 95.88% |

TABLE XXXVII

COMPARISON OF MEAN PAYOFFS OF SIMULATIONS GROUPED BY NO. OF PSNE

| | Simulations | | | | |
|---|---|---|---|---|---|
| Number of PSNE | $\Psi^{2,10,r}_{hom}$ | $\Psi^{2,2,r}_{hom}$ | $\Psi^{2,10,b}_{hom,\varnothing}$ | $\Psi^{2,2,b}_{hom,\varnothing}$ | $\Psi^{2,10,b}_{hom}$ |
| 0 | 0.0032 | 0.0159 | 0.0425 | 0.0573 | 0.0005 |
| 1 | 0.3251 | 0.313 | 0.1248 | 0.1314 | 0.3111 |
| 2 | 0.2352 | 0.2505 | 0.0372 | 0.0366 | 0.257 |
| All | 0.2732 | 0.2678 | 0.0682 | 0.0753 | 0.1895 |

TABLE XXXVIII

CORRELATION OF NNS' OUTPUT AND THE MSNE PRESCRIPTION IN VARIOUS SIMULATIONS

| | Simulations | | | | |
|---|---|---|---|---|---|
| Correlation | $\Psi^{2,10,r}_{hom}$ | $\Psi^{2,2,r}_{hom}$ | $\Psi^{2,10,b}_{hom,\varnothing}$ | $\Psi^{2,2,b}_{hom,\varnothing}$ | $\Psi^{2,10,b}_{hom}$ |
| $\rho_{spearman}$ | 0.1821 | 0.3106 | -0.0067 | 0.0216 | 0.1667 |

### E.1. *Does balancing the training data significantly affect NN behavior?*

As regards the NNs' behavior in terms of obeying dominance principles, running the same simulation with balanced data instead of random data does not lead to any significant changes, as can be seen in Table XXXIII.

TABLE XXXVI.— Comparison of equilibrium selection behavior of NNs in various simulations

| Simulations | Probability of playing types of equilibria | | | | | | |
|---|---|---|---|---|---|---|---|
| | $p(PD)$ | $p(RD)$ | $p(PD\|\sim RD)$ | $p(RD\|\sim PD)$ | $p(PD\&RD)$ | $p(\sim PD\&\sim RD)$ | $RD:PD+RD$ |
| $\Psi^{2,10,r}_{hom}$ | 0.5286 | 0.6869 | 0.0826 | 0.6663 | 0.7259 | 0.1055 | 0.8407 |
| $\Psi^{2,2,r}_{hom}$ | 0.5458 | 0.6843 | 0.0713 | 0.6585 | 0.7311 | 0.1069 | 0.8574 |
| $\Psi^{2,10,b}_{hom,\varnothing}$ | 0.3003 | 0.3127 | 0.214 | 0.2992 | 0.3384 | 0.2299 | 0.5645 |
| $\Psi^{2,2,b}_{hom,\varnothing}$ | 0.3036 | 0.3136 | 0.2135 | 0.2996 | 0.3388 | 0.2296 | 0.5644 |
| $\Psi^{2,10,b}_{hom}$ | 0.5374 | 0.6655 | 0.1365 | 0.6396 | 0.7147 | 0.1102 | 0.7748 |

The probabilities are virtually identical for both simulations with however a slight decline in compliance for the balanced simulation.

In terms of the probability of jointly playing a PSNE again the differences are not very large after balancing the training dataset as exemplified by Table XXXIV. The results are again quite similar in Table XXXV which looks at PSNE play broken down by the number of players with a dominant action. The balanced simulation shows roughly 5% less probability of joint NE play in games where only one player has a dominant action for NNs trained with balanced datasets.

Equilibrium selection behavior does not vary much between the $\Psi_{hom}^{2,10,b}$ and $\Psi_{hom}^{2,10,r}$ simulations. From Table XXXVI it is evident that the largest differences in behavior occur for $p(PD|\sim RD)$ and for $RD : PD+RD$, showing an increased bias towards risk dominant equilibria for the NNs in simulation $\Psi_{hom}^{2,10,r}$.

When examining the average payoffs for each type of game, as defined by the number of PSNE, the differences are again minimal as can be seen in Table XXXVII. The mean payoffs to all games are different simply because of the different composition of the datasets i.e. it consists of a smaller proportion of games with a single PSNE which are the ones leading to the highest payoffs. The correlation of the NNs' behavior to the MSNE prescription is also very similar, see Table XXXVIII.

Concluding, changing the proportion of types of games in the training dataset has not led to any significant changes in the final behavior of the NNs and by implication the learning process during training.

### E.2. *Is there a significant change in behavior when simulations consist only of two agents?*

The two simulations that must be compared to examine this question are the $\Psi_{hom}^{2,2,r}$ and $\Psi_{hom}^{2,10,r}$ simulations. The hypothesis is that when playing against the same opponent it should be easier for agents to coordinate on the same PSNE when games have two PSNE.

As far as complying to dominance principles, there is no significant difference in Table XXXIII between the agents of the two simulations, although in the $\Psi_{hom}^{2,2,r}$ simulation there is slightly less compliance. Joint play of PSNE is again similar with the biggest difference occurring in games with a unique PSNE, where joint PSNE play falls from 95.85% in $\Psi_{hom}^{2,10,r}$ to 89.49% in $\Psi_{hom}^{2,2,r}$. Differences are more pronounced when PSNE play is examined through the prism of the number of players with a dominant action. The $\Psi_{hom}^{2,2,r}$ simulation performs worse in both cases, the largest difference occurs in the case where only one player has a dominant action, falling from 95.01% to 87.42%.

Finally, the data in Table XXXVI does not support the hypothesis that the number of agents in the simulation significantly affects the probability of attaining a risk dominant equilibrium relative to a payoff dominant equilibrium.

### E.3. *Are simple linear neural networks (equivalent to logit regression models) good candidates for modeling human behavior?*

The behavior of linear neural networks in simulations $\Psi^{2,10,b}_{hom,\varnothing}$ and $\Psi^{2,2,b}_{hom,\varnothing}$ in terms of playing dominant actions when they exist is roughly 72%. This performance is much worse than human subjects in the experimental studies reviewed in Section 5.4, which find adherence rates of 90% and above. This is the first piece of evidence that such networks may be too simplistic to model human behavior adequately.

The next inadequacy of linear networks can be traced to the results regarding convergence to PSNE equilibria, especially in games with two PSNE. The probability of these networks playing either of the two PSNE is roughly 50% which is equal to that expected by chance since there are 4 possible joint outcomes in a $2 \times 2$ game. Although humans do have trouble coordinating in such games, experimental evidence finds that they do better than chance. Regarding equilibrium selection, the linear networks show hardly any preference for risk dominant equilibria as $RD : PD + RD$ is roughly equal to 0.56.

Overall, linear networks did not model human subjects well on a number of dimensions as they exhibited less sophisticated behavior than is observed in human experiments.