On Novelty Driven Evolution in Poker

Jessica P. C. Bonson
Faculty of Computer Science
Dalhousie University
Halifax, NS. Canada
Email: jpbonson@gmail.com

Andrew R. McIntyre
Faculty of Computer Science
Dalhousie University
Halifax, NS. Canada
Email: armcnty@cs.dal.ca

Malcolm I. Heywood
Faculty of Computer Science
Dalhousie University
Halifax, NS. Canada
Email: mheywood@cs.dal.ca

***Abstract*—This work asks the question as to whether 'novelty as an objective' is still beneficial under tasks with a lot of ambiguity, such as Poker. Specifically, Poker represents a task in which there is partial information (public and private cards) and stochastic changes in state (what card will be dealt next). In addition, bluffing plays a fundamental role in successful strategies for playing the game. On the face of it, it appears that multiple sources of variation already exist, making the additional provision of novelty as an objective unwarranted. Indeed, most previous work in which agent strategies are evolved with novelty appearing as an explicit objective are not rich in sources of ambiguity. Conversely, the task of learning strategies for playing Poker, even under the 2-player case of heads-up Limit Texas Hold'em, is widely considered to be particularly challenging on account of the multiple sources of uncertainty. We benchmark a form of genetic programming, both with and without (task independent) novelty objectives. It is clear that pursuing behavioural diversity, even under the heads-up Limit Texas Hold'em task is central to learning successful strategies. Benchmarking against static and Bayesian opponents illustrates the capability of the resulting Genetic Programming (GP) agents to bluff and vary their style of play.**

I. Introduction

Poker represents a long standing game of interest for constructing non-person characters (NPC) on account of both the incomplete and stochastic nature of state information [1], [2]. Considerable progress has been made relative to the development of NPC capable of beating strong human opponents, particular when using a combination of game theoretic and exploitive counter-strategies (reviewed below).

In this work, our goal is not to create the strongest possible NPC for Poker, but to use Poker as an environment for asking to what degree novelty as an objective is still of relevance when attempting to develop a range of NPC behaviours from a single run of an evolutionary algorithm. That is to say, there has been sustained interest in the utility of novelty/behavioural diversity as an objective[1] while developing NPC strategies under gaming contexts, e.g., 3-D tic-tac-toe [3], soccer playing agents [4] or Ms. Mac-Man [5]. Conversely, Poker is a task that imparts a lot of ambiguity in the available state information. The basic question we are interested in investigating in this work is whether the explicit promotion of novelty (as a desirable evolutionary trait) is still of fundamental relevance when attempting to evolve NPC behaviours under Evolutionary Computation (EC) for a task with multiple sources of uncertainty.

[1]Either through novelty as an objective, use of multiple performance objectives or a combination of both.

II. Related work

Rubin and Watson identify four general categories of approach as pursued for identifying NPC for playing Poker [2] (where these are often used in combination in practice):

- Game theoretic – focus on achieving optimal play through the construction of a game tree. Given the amount of ambiguity present in the task, a lot of computational/memory resource is invested in parameterizing the game tree (e.g., Monte Carlo rollouts). The recent solution to the heads-up (2 player) Limit Texas Hold'em Poker game utilized 200 computational nodes, each with 24 cores, 32GB RAM and a Tera byte of HD in order to support the division of the game into 110,565 sub-games [6]. The resulting computation took the equivalent of 900 core-years.
- Knowledge based systems – utilize information gained from expert play to describe strategies for NPC play, e.g. [1]. More recently, case based reasoning has been used as the source of information to construct the knowledge of Poker strategies. The approach is potentially more adaptive/scalable as, depending on current state, the case knowledge deployed may change [7], [8].
- Exploitive counter-strategies – emphasize the development of opponent models that are then used to develop an explicitly exploitive strategy of play. When used in combination with the game theoretic approach this represents the state-of-the-art strategy for heads-up Limit Texas Hold'em [6].
- Simulate and learn – combine simulated play against known opponent strategies and/or self play to identify a new NPC strategy through learning. Examples include Bayesian networks, reinforcement learning or neuro-evolution (discussed below). Naturally, such an approach is also sensitive to the quality of experiences encountered during simulation.

The approach adopted here utilizes GP, hence assumes an approach falling into the last category. With this in mind we make the following more detailed observations regarding previous research.

Baker *et al.*, assume the NEAT framework for evolving neural network strategies for Poker NPCs and concentrate

978-1-5090-4240-1/16/$31.00 ©2016 IEEE

on the relative improvement when adopting Bayesian models to characterize opponent behaviour [9]. They also quantify the contribution of support for recurrent connectivity in the evolved networks. All simulations were performed using a simplified one card version of Poker. This builds on an earlier work by the same authors in which four styles of Poker play (loose aggressive/passive, tight aggressive/passive [10]) are modelled and evaluated against an a priori nemesis 'anti-player'. The authors then develop a Bayesian opponent to switch between the relevant 'anti-player' given knowledge of past performance [11].[2] We will later use this model for defining opponents for evolving the NPC players identified through GP. Other researchers have considered evolutionary methods for adapting thresholds determining the point at which basic strategies are switched between in the game of 'Guess It' [13] or under simplified forms of Poker for constructing players with specific strategies [14], [15].

Several researchers have asked what features have most influence on the styles of Poker play identified under evolutionary frameworks for NPC strategy identification. Initially the opponents took the form of static policies [16] whereas later work introduced separate coevolved populations for each NPC [17]. In the latter case, specific attention is given to how to sample 'useful opponents' during evolution in order to minimize pathologies that result in evolving weak NPC strategies.

Nicolai and Hilderman pursued an approach to No-limit Texas Hold'em in which a neural network architecture was evolved to suggest the raise/call/fold decisions as well as the relative amounts bet [18], [19]. The authors assess the relative merits of incorporating more advanced features such as a coevolutionary multi-population formulation and a hall of fame (HoF). Specifically, coevolution was used to mimic an evolutionary arms race between agents sampled from different populations. The HoF archives previously useful strategies in an attempt to reduce the likelihood of forgetting or cycling.

III. Evolving GP Teams

In this work we make use of a previous GP framework for evolving teams of programs, or Symbiotic Bid-based GP (SBB) [20], [21]. Such a framework provides a flexible architecture for promoting the evolution of complex models for classification [21], [22] and reinforcement learning tasks [23], [4]. In the following we highlight some of the properties of the framework before introducing the diversity measures.[3]

A. Properties of SBB

Independent representation of team and program: Teams and programs are represented by two independent populations: Host and Symbiont respectively. Each member of the host population represents a potential team, whereas members of the symbiont population represent candidate programs for appearing within a (host) team. Fitness is only explicitly expressed at the level of (host) team. After each generation the Gap worst performing teams are deleted, and any programs failing to be indexed by at least one team are deleted. Variation operators act on the remaining content generating new individuals by: 1) cloning a team, 2) adding/deleting pointers, 3) cloning one or more symbiont program and then adding/deleting/modifying instructions [20], [21].

Separating context and action: A 'bidding' metaphor is used to explicitly separate the issue of learning a *context* (for an action) and suggesting the action itself [20], [21]. Given the current state from the task domain and a (host) team currently under evaluation, evaluate each *symbiont program* which are a member of this (host) team. The program with maximum output 'wins' the right to suggest its corresponding action. Under reinforcement domains, actions take the form of atomic task specific actions, in the case of Poker three actions are assumed: raise, call, or fold. Each program may only assume a single action. Thus, a team must index programs with at least two different actions, however, the compliment of symbiont programs in any given team represents an evolved emergent property. Different teams may have different numbers of programs, symbiont programs can appear in more than one team, and teams may have multiple programs with the same action. All this freedom in how programs can be deployed by (host) teams provides a wide range of mechanisms for discovering task specific decompositions [21], [4]. Finally, we note that in order to provide the ability to recall previous state (the equivalent of recurrent connections in a neural network) under GP we assume a linear representation (e.g., [24]) in which the content of the registers is retained between consecutive program executions. That is to say, at $t = 0$ (i.e., before the first card is dealt) the register state is initialized (to zero) thereafter, registers retain state between each program execution. Only after the final state is known (e.g., showdown) will the registers be reset.

Inter-host diversity: Despite the use of a teaming metaphor there is no guarantee that a single champion team will emerge that solves all of a task at the end of evolution. As a consequence, mechanisms to encourage inter-host diversity are utilized: implicit diversity maintenance (fitness sharing [3], [21]), multi-objective formulations [5], diversity as an objective [25], or some combination of fitness maximization and novelty [26], [27], [4]. This means that a population of teams with a range of behaviours may emerge that potentially covers the total set of policies necessary to solve a task (e.g., [28]). It is this latter trait that we are particularly interested in under the specific example of the Poker task, i.e. a game of incomplete information with multiple sources of ambiguity opposed to most of the previous research on novelty which tends to assume some form of robotic control with little/no stochastic properties and complete information.

B. Diversity as an objective

Fitness is calculated using a dual objective Pareto optimization (score and diversity); the first will be expressed as

[2] Saund describes a prior work in which more information is made public than available in practice [12].

[3] Several code bases are publicly available http://web.cs.dal.ca/~mheywood/Code/

the score of each team relative to a sample of hands against various opponents (Section IV-E). Diversity will be maintained through the use of the following two diversity mechanisms:

Team complement: expresses diversity in terms of a pairwise comparison in program membership [29]. Thus, the distance between two teams tm_i and tm_j is summarized as the ratio of active programs[4] common to both teams, or

$$dist(tm_i, tm_j) = 1 - \frac{Tm_{active}(tm_i) \cap Tm_{active}(tm_j)}{Tm_{active}(tm_i) \cup Tm_{active}(tm_j)} \quad (1)$$

where $Tm_{active}(tm_x)$ represents the set of active programs in team x. Naturally, such a diversity metric is task independent and explicitly ignores hitchhiking programs.

Behavioural diversity: assumes the concept of a 'normalized compression distance' [26], [29]. That is to say, quantized state-action pairs are recorded for each agent over a set of games. Only the k most unique traces are retained per team. For any pair of teams, the behavioural distance for the two teams quantized state-action pairs is denoted by:

$$dist(h_i, h_k) = NCD(\vec{P}_i, \vec{P}_k) = \frac{Z(\vec{P}_i\vec{P}_k) - min(Z\vec{P}_i, Z\vec{P}_k)}{max(Z\vec{P}_i, Z\vec{P}_k)} \quad (2)$$

where $Z(\vec{P}_i\vec{P}_k)$ is the compressed length of the two profile vectors $\vec{P}_i$ and $\vec{P}_k$ for team i and k respectively. The equation leverages the ability of compression algorithms to filter redundancies in data. For example, if $\vec{P}_i$ and $\vec{P}_k$ are very similar, then $Z(\vec{P}_i\vec{P}_k) \rightarrow Z\vec{P}_i \rightarrow Z\vec{P}_k$, in which case $NCD(\vec{P}_i, \vec{P}_k) \rightarrow 0$; otherwise NCD approaches 1 with increasing difference between vectors. NCD is informative even when comparing vectors that differ in length, which is important because hands will not always complete in a common number of interactions.

IV. Experiments

This work examines the role of opponents as well as novelty mechanisms on the ability to evolve diverse and non-trivial SBB agents capable of playing the full game of heads-up Limit Hold'em Poker. In these experiments, training opponents are drawn from: 1) a static pool, 2) trained Bayesian models, or 3) are the result of self-play. Moreover, both fitness alone and diversity mechanisms are assumed during training. The SBB parameters used in this work are provided by Table I and are common across all three training scenarios. Table II characterizes how cards are drawn to describe games and the types of opponents encountered. In the following we detail how these parameters are arrived at.

A. Poker Task: heads-up Limit Texas Hold'em

In this section we briefly describe the heads-up Limit Hold'em game and basic rules of the game. The game of *heads-up* Texas Hold'em is a two-player card game where each player wagers (or not, as players are free to exit the game, or 'fold', during any betting round) on the prospect that they hold the highest ranking five-card Poker hand. Each player must use the best possible combination of five visible (community) cards and two private (so-called 'hole') cards to make up their five-card Poker hand. Community cards are upturned (face up) and considered public information as opposed to the hole cards of each player, which are hidden information from the perspective of the opponents. 'Limit' is a version of Texas Hold'em where betting rounds are fixed with respect to bet size (some number of 'chips') and total number of bets that can be made and 'heads-up' is the designation given to the two-player incarnation of the game.

The game (or 'hand') begins with users paying forced or 'blind' bets that are fixed relative to which player is considered to be playing in the 'dealer' position. A token (a.k.a., 'button') is circulated in order to designate this player for each game so that every player will eventually pay the blind bets before the cards are dealt. The blind bets are used to incentivize the betting action so that some amount is always at stake, even before any cards are dealt. With the blind bets in the pot, the dealer proceeds to deal two hole cards to each player and a betting round (known as 'pre-flop') takes place. A set of three community cards are dealt (referred to as the 'flop') and are public information for which players can begin to formulate an estimate of their hand strength and potential. Moreover, another betting round ensues where players might infer something about the strength of their opponents hand. An additional public card is dealt ('turn') and yet another betting

[4]An active program is one that wins at least one bidding round during evaluation across multiple training games.

TABLE I
General SBB Parameters for Training Poker Agents. Instructions have the general form $R[x] = R[x]\langle op\rangle R[y]$ or $R[x] = \langle op\rangle R[y]$ depending on the operator arity. x and y represent register references and a mode bit may toggle $R[y]$ to index a state attribute. Relational instructions have the form: IF $R[x]\langle op\rangle R[y]$ THEN $\langle instruction\rangle$]

General Parameter	Value
Runs	25
Generations	300
Teams	100
Team Replacement Rate	0.5
hand Replacement Rate	0.2
Selection Type	Uniform
Team Size	Min: 2, Max: 16
Program Size	Min: 5, Max: 40
Total Registers	5
Operators	$\langle op\rangle \in \{+, -, /, *, \ln, \exp, \cos\}$
Relational	$<, \geq$
Reproduction	Mutation
Team Mutation	Program: Add / Del: 0.7, Mutate: 0.2
Instructions	Add / Del: 0.5, Change: 1.0, Action: 0.1

TABLE II
Hand and Opponent Parameters for Training Poker Agents

Hand / Opponent Parameters	Value
Total Hands	600
Static Opponents (LA, LP, TA, TP)	4 (one of ea.)
Bayesian Opponents	4
HoF Opponents (training only)	0-2

round ensues. The fifth and final card is dealt ('river') and the last round of betting takes place. Assuming no players fold on the river round, the players then enter the showdown phase, revealing their hole cards. The player with the highest ranking five-card Poker hand wins the chips in the pot and pots are split in the event of a tie. More details on the game of Limit Texas Hold'em are available from multiple sources, e.g. [30].

B. Training of SBB Poker Agents

We consider three training scenarios and provide a basis for the learning context with the reinforcement style SBB learning environment configured for the various training scenarios [23], [4]. At each training epoch, a set of 600 hole cards are sampled such that they have similar strength across nine pre-defined hand groupings according to hand strength [1]. Such an approach is taken on account that hands drawn randomly will represent weak hands in the vast majority of cases. Under such a condition it is most likely that a learning agent will just 'fold', thus not learning anything.

Specifically, hands are broadly categorized across three groups as weak, intermediate and strong. Hands are deployed evenly across both agent and opponent starting hands (for a total of nine groups). Whereas an unbalanced dealing refers to the traditional dealing of starting hands in which a real-world game of Texas Hold'em Poker result in approximately 60% of all hands being considered weak, 30% intermediate and 10% strong. For the balanced distribution, the hands are equally distributed between the opponents, so that if there are 4 opponents and 72 hands, each opponent will participate in 2 hands for each of the 9 types of hand balance. The three training configurations for SBB are described as follows:

1) no diversity, trained on static opponent pool and HoF;
2) with diversity, trained on static opponent pool and HoF;
3) with diversity, trained on static opponent pool, HoF and Bayesian opponents.

There are a total of 14 inputs, compatible with that of the standard Limit Texas Hold'em Poker environment established by the Annual Computer Poker Competition (ACPC)[5], normalized between 0.0 and 10.0. The inputs are divided in two groups, for inputs about the hand (Table III) and inputs modeling the opponent (Table IV). Effective Potential is composed of hand equity (for the pre-flop), hand potential (for the flop and turn), and hand strength (for the river) [1]. Equity in the pre-flop is used as a way to deal with the infeasibility of simulating all the next possible hands, and hand strength is used in the river since it estimates the potential for that round.

C. Opponents

Three types of opponents are deployed to play as SBB adversaries in the heads-up Limit Hold'em task: Static, Bayesian, and HoF. The Static pool consists of one of each of the four classical opponent types (LP, LA, TP, TA) [10], [11]. For each classical opponent type we characterize the play type as {loose, tight} and either {passive, aggressive} [10] where each pair is controlled by α in the case of the former and β in the case of the latter. These parameter-controlled opponents were defined using the model from the work of Baker et al., [11], [9] using the hand strength as the winning probability and the α and β parameters. The α parameter defines the threshold of hand strength so that the player enters the hand (or folds) whereas β defines how passively or aggressively the player will take action (call or raise). Opponent types are described in more details below.

[5]https://github.com/jmasha/acpc_poker_client

TABLE III
GAME STATE INPUTS ABOUT HAND

Input	Description
Hand Strength	Current best hand rank [1]
Effective Potential	Hand equity, potential [1], and strength
Pot Odds	Hand value relative to pot size [1]
Betting Position	First or second to bet, for current round
Betting Round	Pre-flop, Flop, Turn, River

TABLE IV
GAME STATE INPUTS ABOUT OPPONENT

Input	Description
Last Action	The last action in current hand
Long-term Aggressiveness	Opponent aggression over time [19]
Short-term Aggressiveness	Recent opponent aggression [19]
Hand Aggressiveness	Opponent aggression, current hand
Tight / Loose	How many hands the opponent played
Passive / Aggressive	Ratio of calls to raises
Bluffing	Freq. of raises with weak hands at river
Chips	Current chips won / lost vs. opponent
Self Short-term Aggres.	Player's aggression against opponent [19]

LP Loose passive opponents are characterized by low requirements for hole card strength ($\alpha = 0.2$) and less likely to raise during the ensuing betting rounds ($\beta = 0.8$).

LA Loose aggressive opponents are know as the 'wild man' strategy, characterized by low requirements for hole card strength ($\alpha = 0.2$) and more likely to become the aggressor by raising during betting rounds ($\beta = 0.4$) despite playing a potentially volatile range of hand strengths.

TP Tight passive combines the high requirements for starting hand strength ($\alpha = 0.8$) with passivity in the betting rounds ($\beta = 0.95$).

TA Tight aggressive opponents are typically seen as a relatively dangerous combination between hand selection requirements ($\alpha = 0.8$) with aggressiveness in the betting rounds ($\beta = 0.85$).

Bayes The Bayesian opponent was implemented following the model of Baker et al., [11], [9] with a few adaptations to tune the parameters for Texas Hold'em Poker. In the original model, four players are trained so that α and β values are optimized against the four classical opponents. Finally a Bayesian model is used to define which of the four styles the opponent is playing, based on their actions and on predefined action probabilities per style. The final step is to

execute an action from the player that is stronger against the estimated style of the opponent.

HoF Champion agents are retained for training via a Pareto archiving mechanism that considers both fitness and NCD diversity as dual objectives. Archiving proceeds such that a maximum of two HoF opponents are eventually applied during the training process.

D. Testing of SBB Poker Agents

Testing is conducted with respect to two configurations, namely: 1) Balanced distribution of hands, and 2) Unbalanced distribution of hands. The balanced distribution represented the distribution of card hand strengths (weak, intermediate, strong) employed during training, whereas the unbalanced distribution represents the distribution that would be encountered under a practical setting (60%, 30%, 10%). In each case, we considered Static opponents and Bayesian opponents.

E. Fitness

The fitness function for individual SBB teams is calculated as the number of chips won per hand, considering the maximum number of chips that can be theoretically won or lost. A fitness of 0 indicates that all possible chips were lost (i.e., raised in all their opportunities but lost at the showdown), 0.5 means they neither lost or won chips (i.e., break even), and 1.0 means they raised every time and won the showdown. Since the Poker game implemented is the full game with 4 betting rounds, small bet of 10 and big bet of 20, the maximum amount of chips a player can either win or lose per hand is 240. So the fitness [0.0, 1.0] is a mapping from [-240, +240]. As an example, a player that obtained a mean fitness of 0.6 across its matches won an average of 48 chips per hand. The same function is employed in all testing.

Fitness and the diversity measures of Section III-B are combined through the following two step process [27]:

1) Stochastically select one diversity measure at the beginning of each generation. This minimizes the computational overhead of supporting multiple diversity measures.
2) Rank the combined fitness and diversity as two objectives using Pareto dominance. This implies that the Pareto rank of individuals in the (2-dimensional) objective space establishes their relative quality, hence fitness and diversity represent equally important 'objectives'.

V. Results

Test result distributions are collected for 1260 hands against each opponent to demonstrate the overall hand performance (scores – ranging from 0 to 1260, i.e., sum of fitness values across hands) to illustrate the average number of chips won during the test scenario, with 0 indicating maximum loss to 1260 indicating maximum won across all hands played.

Post-training, the entire performance of the population is ranked according to median score performance of each individual (descending) and the combined 'cumulative' (ascending) scores. That is to say, given individual i then the cumulative score reflects the score assuming an 'oracle' may choose the best individual to play the hand from the set $\{1, ...i\}$. The cumulative curve will never be worse than the individual-wise curve, however, if the cumulative curve is significantly better, this indicates that individuals have identified non-overlapping strategies that complement each other.

Both 'random' and break-even ('draw') lines are provided in each performance chart to illustrate where a random act (no environment information) performs against opponent curves and where each team would have to perform to simply break even (return of zero chips), respectively. Also, for the hands against static opponents there is a line corresponding to the performance of the Bayesian opponent.

Additionally, we compared the final results between the teams trained and not trained against the Bayesian opponent. A (non-parametric) Mann-Whitney U test with significance level less than 0.05 was applied to test the distributions of the individual and cumulative performances for the 25 runs in all the test scenarios. The only scenarios that were statistically different represent those against the Bayesian opponent, with balanced hand distributions. With this in mind, we first introduce the base case of no diversity, and then comment on performance against the static LA and then Bayesian opponent (space precluding reporting of all possible combinations).

A. No Diversity

A control experiment is conducted in which SBB is trained without any diversity mechanisms (as described in section IV-E). Individual and cumulative performance results against the static LA opponent across 1260 unbalanced test hands are summarized in Figure 1, and against the Bayesian opponent in Figure 2. The performance distributions are nearly flat across both individual-wise and cumulative curves, indicating not only poor performance in teams, but also a distinct lack of diversity between teams since cumulative recombining of teams was not able to return an improvement in score. Moreover, the cumulative-wise score are just barely better than the Bayesian score against the LA opponent. We note, however, that the best performing no diversity SBB team returns a score that is significantly better than random. The case for balanced test hands returned results that were largely similar, however with best scoring no diversity SBB teams performing worse (just breaking-even) than the unbalanced cases.

B. Diversity SBB vs. Static opponents

Figure 3 summarizes scores under test games for SBB evolved with diversity maintenance versus the static LA opponent. The single best individual already outperforms the opponent (i.e., score above the 'draw' line), and a strong cumulative curve is also apparent with different individuals clearly being effective under different hands. In fact, in all test cases, for both balanced and unbalanced against all the 4 static opponents, the best individual team outperforms the static opponents, along with strong cumulative curves. Another

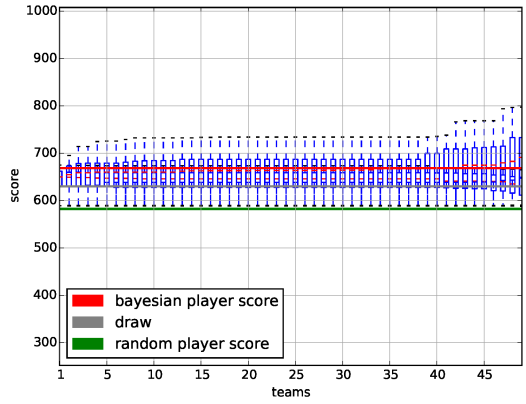

Fig. 1. No diversity SBB vs. LA opponent, for 1260 unbalanced hands.

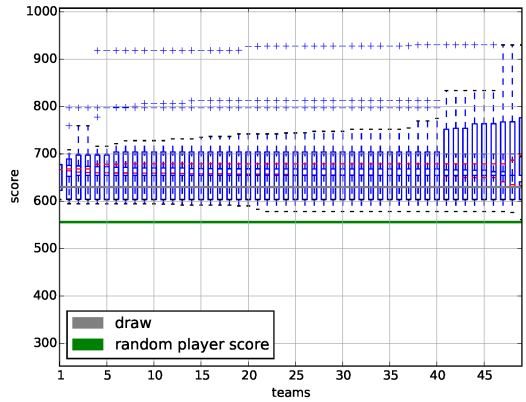

Fig. 2. No diversity SBB vs. Bayesian opponent, for 1260 unbalanced hands.

clear observation in the chart is that the Bayesian opponent tends to win more chips than the single best SBB individual when playing against the same static opponents. However, these results do not characterize performance when against each other (see the next section).

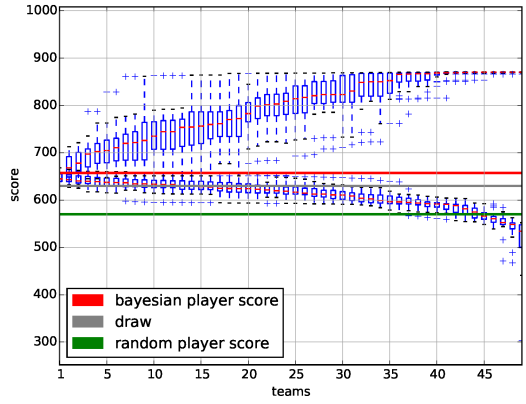

Fig. 3. Diversity SBB vs. LA opponent, for 1260 unbalanced hands.

C. Diversity SBB vs. Bayesian

Figure 4 shows the outcome of SBB with diversity playing 1260 hands against the Bayesian opponent. Indeed the top 32 (out of 50) individuals performed better than the Bayesian model and the cumulative performance climbs appreciably within the first 75% of the population. Moreover, this property is common both for the teams that were/weren't trained against the Bayesian opponent, so diversity SBB is consistently shown to outperform the Bayesian opponent. Clearly, the maintenance of diversity is the key to this property. Additionally, the diversity measures used are not task specific (Eqs. (1) and (2)), thus do not rely on intuition regarding what might be an appropriate task specific measure of diversity.

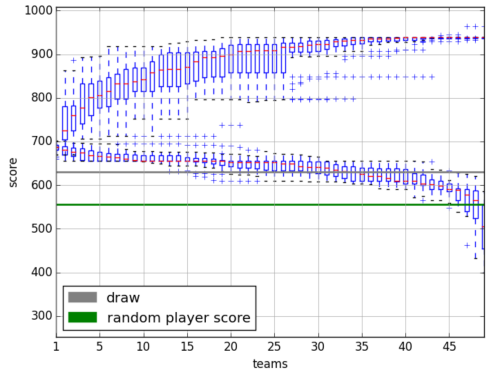

Fig. 4. Diversity SBB vs. Bayesian opponent, for 1260 unbalanced hands.

D. Behavioral Properties of SBB agents

The behavioral properties of trained SBB agents are provided in the following plots and include: win rate, hands played, bluff frequency, and aggression. In general it is clear that the real-world Poker hand distributions (unbalanced hands) lead diversity SBB to behave more aggressively (and with more success) than in the balanced scenarios. All the properties were computed against the Bayesian opponent, using 1260 balanced and 1260 unbalanced hands.

1) Hands Played and Won: In terms of play (Figure 5), we note that more hands are played in the unbalanced test scenario. This is intuitively necessary given the broader range of hands. Specifically, the best scoring teams are able to play around 60-80% of hands for their maximum score in the unbalanced tests. Balanced hand performance peaks near 50% of hands played, which is another intuitive result given that the hole cards are very close in strength. Winning outright, however, is another matter (Figure 6). Here we found that the best agents were typically only winning between 35 and 40% of the hands overall. Moreover, a plot of cumulative potential as a result of winning alone against all opponents (sample against LA opponent in Figure 7) does not increase with the addition of teams. This suggests that the best individual teams are already winning the hands that can be typically won (on average) and that the additional behaviors are the means by which combinations of teams are able to exploit opponents and achieve better cumulative results.

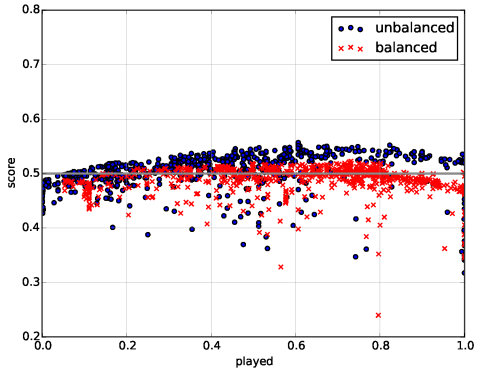

Fig. 5. Play rates vs Score of SBB agents.

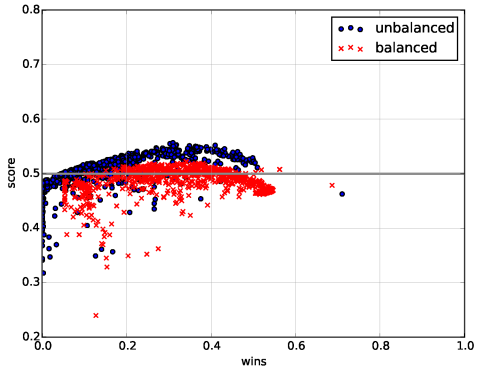

Fig. 6. Win rates vs Score of SBB agents.

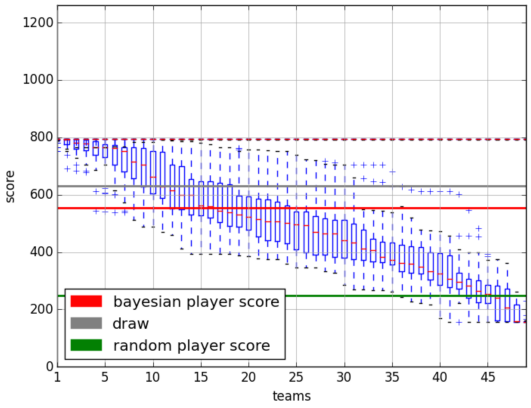

Fig. 7. Diversity SBB vs. Bayesian opponent, ranked by hands won for 1260 unbalanced hands.

2) Bluffing and Aggression: In both the cases of bluffing (choosing to raise on the river, despite a weak hand) and aggression (stronger tendency to raise, then to call, and then to fold), the SBB agents are seen to increase the levels of these behaviors in the unbalanced test scenarios (Figures 8 and 9). While this is relatively intuitive, given the greater likelihood of hand differential in real-world hand distributions, it is interesting to note the success of bluffing and aggression in the unbalanced test case. In terms of bluffing, the utility falls off at around 10% in balanced cases whereas bluffing behavior continues to be useful up to approximately 50% of the time in the unbalanced testing scenario. It is clear that bluffing behavior in the most successful teams is deployed about 30% of the time against the Bayesian opponent (Figure 8). Given that the cards dealt under the unbalanced scenarios contain more weak hands (and vice versa under the balanced) this appears to indicate that bluffing/aggressive play is being used under the weaker hands to hide the true strength of the private cards. It is intuitive that the converse appears under the balanced scenarios where there are stronger hands.

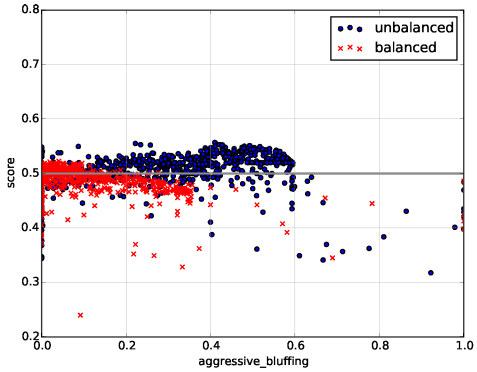

Fig. 8. Bluffing rates vs Score of SBB agents.

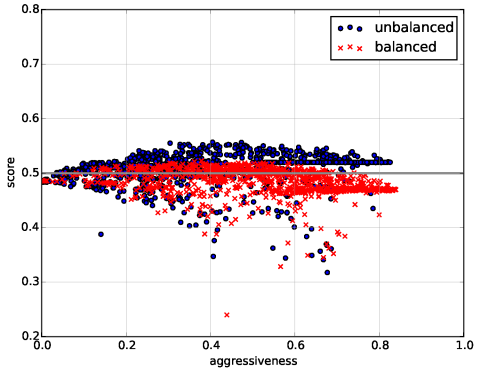

Fig. 9. Aggressiveness vs Score of SBB agents.

E. Most Used Inputs

The Table V shows the most used inputs, by the percentage of teams that had at least one program that used it. This data was obtained at the end of the training where the teams trained against the Bayesian opponent. Perhaps surprisingly, three of the five most used attributes are opponents' inputs, while the top two are environment inputs. A few of the inputs were nearly redundant, but they were available so the teams could 'choose' which ones worked the best. So it is interesting to see that more teams used 'effective potential' than 'hand strength', and that inputs that modeled the opponent's aggressiveness on short-term were more employed than the ones for long-term behavior.

TABLE V
SBB TEAM MOST USED INPUTS

#	Input	% Usage
1	pot odds	82.7
2	effective potential	80.1
3	opp passive/aggressive	73.5
4	opp hand agressiveness	72.0
5	opp short-term agressiveness	71.9
6	hand strength	70.9
7	opp bluffing	69.9
8	round	68.8
9	self short-term agressiveness	68.8
10	chips	67.4
11	opp tight/loose	65.0
12	betting position	64.5
13	opp agressiveness	64.2
14	opp last action	63.8

VI. CONCLUSION

The contribution of novelty as an objective is investigated under the highly ambiguous task of evolving strategies for playing heads-up Limit Texas Hold'em Poker. It is demonstrated that supporting novelty in addition to the underlying performance objective of maximizing chips won, represents a significant factor. Additional 'must have' properties include ensuring that training is performed against a balanced sampling of hole card strengths and including a cross section of opponent capabilities. The resulting Poker strategies are capable of better play than the opponents and appear to index attributes to support temporal properties of the game. Moreover, we note that the novelty objectives employed are not task specific, indeed they were first demonstrated under the Keepaway soccer task [29].

In terms of future work, the cumulative curves hint that there is the potential for deploying some subset of the Poker playing agents collectively. However, additional research would be necessary to identify an efficient mechanism for doing so. Current state-of-the-art relies on extensive self-play between different policies to construct game trees capable of resolving this issue [6].

ACKNOWLEDGMENT

The authors are supported by the NSERC CRD program.

REFERENCES

[1] D. Billings, A. Davidson, J. Schaeffer, and D. Szafron, "The challenge of poker," *Artificial Intelligence*, vol. 134, pp. 201–240, 2002.

[2] J. Rubin and I. Watson, "Computer poker: A reivew," *Artificial Intelligence*, vol. 175, pp. 958–987, 2011.

[3] C. D. Rosin and R. K. Belew, "New methods for competitive coevolution," *Evolutionary Computation*, vol. 5, no. 1, pp. 1–29, 1998.

[4] S. Kelly and M. I. Heywood, "On diversity, teaming, and hierarchical policies: Observations from the Keepaway soccer task," in *European Conference on Genetic Programming*, ser. LNCS, vol. 8599, 2014, pp. 75–86.

[5] J. Schrum and R. Miikkulainen, "Evolving multimodal behavior with modular neural networks in Ms. Pac-Man," in *ACM Genetic and Evolutionary Computation Conference*, 2014, pp. 325–332.

[6] M. Bowling, N. Burch, M. Johanson, and O. Tammelin, "Heads-up limit hold-em poker is solved," *Science*, vol. 347, no. 6218, pp. 145–151, 2015.

[7] I. Watson, S. Lee, J. Rubin, and S. Wender, "Improving a case-based Texas Hold'em Poker bot," in *IEEE Symposium on Computational Intelligence and Games*, 2008, pp. 350–356.

[8] J. Rubin and I. Watson, "Successful performance via decision generalization in N Limit Texas Hold'em," in *International Conference on Case Based Reasoning*, ser. LNAI, vol. 6880, 2011, pp. 467–481.

[9] R. J. S. Baker, P. I. Cowling, T. W. G. Randall, and P. Jiang, "Can opponent models aid Poker player evolution?" in *IEEE Symposium on Computational Intelligence and Games*, 2008, pp. 23–30.

[10] K. Burns, "Style in Poker," in *IEEE Symposium on Computational Intelligence and Games*, 2006, pp. 257–264.

[11] R. J. S. Baker and P. I. Cowling, "Bayesian opponent modeling in a simple Poker environment," in *IEEE Symposium on Computational Intelligence and Games*, 2007, pp. 125–131.

[12] E. Saund, "Capturing the information conveyed by opponents' betting behavior in Poker," in *IEEE Symposium on Computational Intelligence and Games*, 2006, pp. 126–133.

[13] A. Di Pietro, L. Marone, and L. While, "A comparison of different adaptive learning techniques for opponent modelling in the game of Guess It," in *IEEE Symposium on Computational Intelligence and Games*, 2006, pp. 173–180.

[14] L. Barone and L. While, "Evolving adaptive play for simplified Poker," in *IEEE Congress on Evolutionary Computation*, 1998, pp. 108–113.

[15] ——, "Adaptive learning for Poker," in *Proceedings of the Genetic and Evolutionary Computation Conference*. Morgan Kaufmann, 2000, pp. 566–573.

[16] R. G. Carter and J. Levine, "An investigation into tournament Poker strategy using evolutionary algorithms," in *IEEE Symposium on Computational Intelligence and Games*, 2007, pp. 117–124.

[17] T. Thompson, J. Levine, and R. Wotherspoon, "Evolution of counter-strategies: Application of coevolution to Texas Hold'em Poker," in *IEEE Symposium on Computational Intelligence and Games*, 2008, pp. 16–22.

[18] G. Nicolai and R. J. Hilderman, "No-limit Texas Hold'em Poker agents created with evolutionary neural networks," in *IEEE Symposium on Computational Intelligence and Games*, 2009, pp. 125–131.

[19] ——, "Countering evolutionary forgetting in no-limit Texas Hold'em Poker agents," in *Computational Intelligence*, ser. SCI, K. Madani et al., Ed. Springer, 2012, vol. 399, pp. 31–48.

[20] P. Lichodzijewski and M. I. Heywood, "Managing team-based problem solving with symbiotic bid-based genetic programming," in *Proceedings of the ACM Genetic and Evolutionary Computation Conference*, 2008, pp. 863–870.

[21] ——, "Symbiosis, complexification and simplicity under GP," in *Proceedings of the ACM Genetic and Evolutionary Computation Conference*, 2010, pp. 853–860.

[22] A. Vahdat, J. Morgan, A. R. McIntyre, M. I. Heywood, and N. Zincir-Heywood, "Evolving GP classifiers for streaming data tasks with concept change and label budgets: A benchmarking study," in *Handbook of Genetic Programming Applications*, A. H. G. et al., Ed. Springer, 2015, ch. 18, pp. 451–480.

[23] J. A. Doucette, P. Lichodzijewski, and M. I. Heywood, "Hierarchical task decomposition through symbiosis in reinforcement learning," in *ACM Genetic and Evolutionary Computation Conference*, 2012, pp. 97–104.

[24] M. Brameier and W. Banzhaf, *Linear Genetic Programming*. Springer, 2007.

[25] J. Lehman and K. O. Stanley, "Abandoning objectives: Evolution through the search for novelty alone," *Evolutionary Computation*, vol. 19, no. 2, pp. 189–223, 2011.

[26] F. Gomez, "Sustaining diversity using behavioral information distance," in *Proceedings of the ACM Genetic and Evolutionary Computation Conference*, 2009, pp. 113–120.

[27] S. Doncieux and J.-B. Mouret, "Behavioral diversity with multiple behavioral distances," in *IEEE Congress on Evolutionary Computation*, 2013, pp. 1427–1434.

[28] S. Y. Chong, P. Tiňo, and X. Yao, "Relationship between generalization and diversity in coevolutionary learning," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 3, pp. 214–232, 2009.

[29] S. Kelly and M. I. Heywood, "Genotypic versus behavioural diversity for teams of programs under the 4-v-3 keepaway soccer task," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014, pp. 3110–3111.

[30] D. Sklansky, *The Theory of Poker: a Professional Poker Player Teaches You How To Think Like One*. Two Plus Two Publishing, 1999.